Biochemical analysis of MBD1

Matthew James Lyst

Thesis presented for the degree of Doctor of Philosophy

Wellcome Trust Centre for Cell Biology

University of Edinburgh

2009

Declaration

I declare that this thesis was composed by myself, and that the research presented is my own except where otherwise stated. This work has not been submitted for any other degree or professional qualification.

Matthew J. Lyst

2009

Acknowledgements

First of all I thank my supervisor, Irina, for taking me on as a PhD student, and for her guidance over the last three and a half years. I also thank all members of her laboratory for making my time there so enjoyable. In particular Kevin Myant for helping me with the language, Jose de las Heras for the regular Friday morning pantomime, Katrina Gordon for telling us off when we needed to be, Cara Merusi for the tea and Thomas Clouaire for his boundless cynicism. I also thank everyone in the laboratory of Adrian Bird for the banter, help and advice, and many stimulating discussions.

I am also grateful to the following for gifts of reagents: Bill Earnshaw's laboratory for plasmids encoding various affinity tags, Aimee Deaton for pCMV10 3xFLAG Mbd2, and Klaus Hansen and Patrick Varga-Weisz for antibodies. I thank Juri Rappsilber and Flavia Alves as well as Axel Imhof and people in his laboratory for carrying out mass spectrometry. I also gratefully acknowledge the contribution of Alex Tuck to the work on the potential self association of MBD1.

Finally I am thankful for the friends I have shared my time in Scotland with, in particular my flatmates Oyinlola Oyebode and Ruth Corrigan, as well as everyone at Sporting ICAPB, the most tremendously brilliantly mediocre football team in the whole of Edinburgh.

The work in Dr Irina Stancheva's laboratory is funded by Cancer Research UK and the Wellcome Trust. I am grateful to the BBSRC for the studentship I received.

Table of Contents

Figure List

Abstract

Methylation of cytosines within CpG dinucleotides is a feature of vertebrate DNA. The precise role of DNA methylation is unknown to date, although it has been implicated in several processes relating to transcriptional regulation. One approach to study DNA methylation is the characterization of proteins that bind specifically to methylated DNA. One such family of proteins is the methyl-CpG binding domain (MBD) containing family and MBD1 is a member of this family. MBD1 is implicated in transcriptional repression and various mechanisms by which it might bring about gene silencing have been proposed. These are mainly based on studies reporting interactions between MBD1 and various proteins that regulate chromatin structure. Also MBD1 function can be modified by PIAS proteins, which stimulate its conjugation to SUMO (small ubiquitin-like modifier).The original aim of this work was to address two questions about MBD1:

(1) Does MBD1 form part of a stable complex with other factors, and if so, what are the identities of the other components? Purification of MBD1 revealed the presence of no stably bound interacting proteins. However, some evidence indicates MBD1 may interact with itself and form dimers, a finding which impacts on many aspects of the function of MBD1. Also a proteomics screen for transient interaction partners identified candidate binding partners for MBD1 and the related protein MeCP2, which may throw light on the function of these proteins.

(2) Are there any activities which regulate MBD1 function by the removal of SUMO from this protein? No activities capable of removing SUMO from native MBD1 were found but it was demonstrated that this modification leads to the destabilization of MBD1 *in vitro*. The relevance of this finding *in vivo* is yet to be determined.

Abbreviations

Å (Angstrom)

APS (ammonium persulfate)

ATP (adenosine triphosphate)

BSA (bovine serum albumin)

DAPI (4',6-diamidino-2-phenylindole)

DMSO (Dimethyl sulfoxide)

dNTP (deoxyribonucleoside triphosphate)

DNA (deoxyribonucleic acid)

DTT (dithiothreitol)

EDTA (ethylenediaminetetraacetic acid)

EMSA (electrophoretic mobility shift assay)

EST (expressed sequence tag)

g (relative centrifugal force)

GFP (green fluorescent protein)

GST (glutathione S-transferase)

h (hour)

HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)

HRP (horseradish peroxidase)

IAP (intra-cisternal A particle)

IPTG (isopropyl β-D-1-thiogalactopyranoside)

kDa (kilodalton)

M (molar)

MBD (methyl-CpG binding domain)

min (minutes)

NEM (N-ethylmaleimide)

NMR (nuclear magnetic resonance)

PAGE (polyacrylamide gel electrophoresis)

PBS (phosphate buffered saline)

PCR (polymerase chain reaction)

pH ($-\log_{10}[H^+]$)

PIPES (piperazine-N,N'-bis(2-ethanesulfonic acid)

PMSF (phenylmethanesulphonylfluoride)

RNA (ribonucleic acid)

SAM (s-adenosyl-L-methionine)

SBP (streptavidin binding peptide)

SDS (sodium dodecyl sulphate)

SUMO (small ubiquitin-like modifier)

TAE (Tris-acetate EDTA)

TBS (Tris buffered saline)

TEMED (tetramethylethylenediamine)

TRD (transcription repression domain)

Tris (2-amino-2(hydroxymethyl)-1,3-propanediol)

TSA (trichostatin A)

UV (ultraviolet)

V (volts)

˚C (degree centigrade)

1. Chapter One - Introduction

1.1 The distribution of DNA methylation

1.1.1 DNA molecules as an information repository

The heritable information required for the development and function of cellular life is carried in the form of DNA molecules (Avery *et al*, 1944). The information content of DNA is embodied in the sequence of bases (adenine, thymine, cytosine and guanine) in each polynucleotide chain. The double helical structure of DNA with complementary base-pairing of adenine to thymine and cytosine to guanine between anti-parallel polynucleotide strands reveals how DNA molecules can be copied, and thus how information can be transmitted through cell divisions (Watson and Crick, 1953). In addition to the heritable nucleotide sequence, extra information can be held in DNA molecules due to the possibility of covalent modifications of the bases, in particular methylation of the carbon 5 or the amine at the 4 position of cytosine, and also the amine at the 6 position adenine (Figure 1.1).

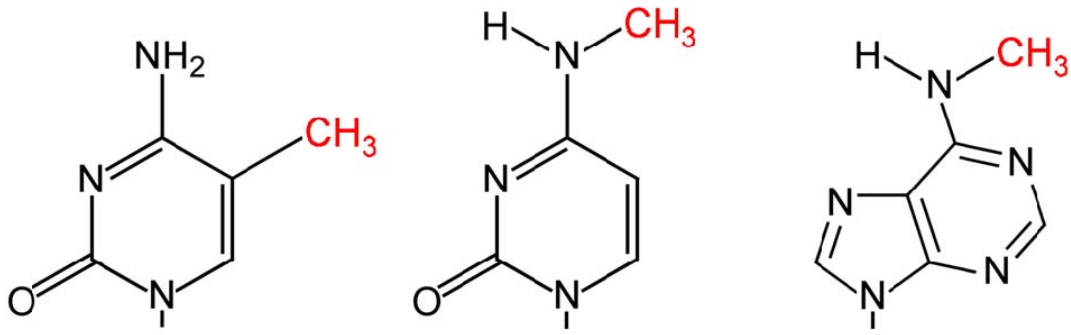

Figure 1.1 – Methylated bases found in DNA

Methylation can occur on the carbon 5 or the amine at the 4 position of cytosine as well as the amine at the 6 position of adenine. The added methyl groups are shown in red.

1.1.2 Bacterial DNA methylation and restriction-modification systems

In prokaryotes DNA methylation is a component of the 'restriction-modification' systems that protect the host from infection by bacteriophages. The host cell expresses DNA methyltransferase enzymes that chemically modify its own DNA as well as restriction endonucleases which specifically cleave DNA whose modification status is different to that of the host. Thus bacteriophage DNA can be recognised as foreign and destroyed by the cell (Wilson and Murray, 1991). Other functions of bacterial DNA methylation have been proposed including roles in DNA replication, repair and transcriptional regulation (Palmer and Marinus, 1994). Eukaryotic cells seem to lack restriction-modification systems but nevertheless many species still contain cytosine methylated at the carbon 5 position in their DNA. This suggests that an alternative function for DNA methylation exists in these organisms.

1.1.3 The distribution of 5-methylcytosine across the eukaryotes

In contrast to the situation in prokaryotes, methylation of eukaryotic genomes has been found exclusively at the carbon 5 position of cytosine. However, perhaps surprisingly for a covalent modification of the very molecule that stores the hereditary information of life, the presence and levels of 5-methylcytosine in eukaryotic genomes appear not to have evolved parsimoniously, and they can vary greatly from one organism to another.

At one extreme organisms including the nematode worm *Caenorhabditis elegans* and the budding yeast *Saccharomyces cerevisiae* both lack detectable cytosine methylation and their genomes encode no proteins closely related to known DNA methyltransferase enzymes (Simpson *et al*, 1986; Proffitt *et al*, 1984; Goll and Bestor, 2005). At the other, some plants such as the tobacco plant *Nicotiana tabacum* methylate up to a third of the cytosine in their DNA (Montero *et al*, 1992). Between these two is a spectrum where evolutionary relationships between various eukaryotes often do not reflect the levels of

methylation they are observed to bear. Adult DNA from the fruit fly *Drosophila melanogaster* contains only trace amounts of 5-methylcytosine whereas DNA from the embryos comprises approximately 0.4% methylated cytosine predominantly in the context of CpT dinucleotides (Lyko *et al*, 2000). This situation is slightly unusual as in many other organisms the DNA methylation level does not depend so strongly on the tissue or life cycle stage examined (Bird and Taggart, 1980). 5-methylcytosine is also readily detectable in adults of the honey bee *Apis mellifera* and other social Hymenoptera (Wang *et al*, 2006; Kronforst *et al*, 2008), and so together with the fruit fly this is an example of how even organisms within the same phylogenetic class can show very different levels of DNA methylation. Fungi also show diverse amounts of methylation. Although *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* are devoid of the modification, this is not the case in *Ascobolus immersus* (Goyon *et al*, 1996) or *Neurospora crassa*, the latter of which methylates 2-3% of its cytosine (Foss *et al*, 1993; Selker *et al*, 2003). Plants too show great variance in the abundance of genomic 5-methylcytosine between species. In contrast to the high methylation levels seen in the tobacco plant, the mustard weed *Arabidopsis thaliana* methylates only 7% of the cytosines available in its genome (Lister *et al*, 2008). An early study found cytosine methylation in all vertebrates examined and its contribution there to genomic DNA from somatic tissues ranges from just under 1% to 2% depending on the species in question (Vanyushin *et al*, 1970). Later work confirmed the presence of methylation in vertebrates and also found the modification, but to a lesser degree, in all non-arthropod invertebrates tested (Bird and Taggart, 1980).

As well as the notable differences in the presence and abundance of DNA methylation between different eukaryotes, the distribution of this modification throughout different compartments of the genome also varies greatly from one organism to another. These methylation patterns are likely to reflect the different functions of DNA methylation across species and will be the subject of section 1.1.5.

1.1.4 Symmetrical methylation of CpG dinucleotides

Early work on cytosine methylation in eukaryotes focussed on describing how this modification is distributed in genomes. Analysis of dinucleotides isolated from a DNAse digest of calf thymus DNA found methylated cytosine only in the context of CpG dinucleotides (Sinsheimer, 1955). A similar situation was subsequently found in the sea urchin with over 90% of its genomic 5-methylcytosine being associated with the sequence CpG (Grippo *et al*, 1968). A more recent nearest neighbour analysis showed that the mean contribution across various mouse tissues (excluding embryonic stem cells) of CpG methylation to total cytosine methylation was 93% (Ramsahoye *et al*, 2000). A somewhat different situation is found in plants with SOLEXA sequencing of bisulphite treated DNA (Frommer *et al*, 1992) revealing 45% of the methylated cytosines in *Arabidopsis thaliana* to occur in sequence contexts other than CpG (Lister *et al*, 2008).

It was subsequently demonstrated that the vast majority of CpG dinucleotides are either symmetrically methylated on both strands or not methylated at all, with hemimethylated sites being rare. An initial line of evidence in support of this was the fact that somatic ^{32}P-labelled Xenopus rDNA retains its resistance to digestion by the methylation-sensitive restriction enzymes HpaII and HhaI even after reassociation with an excess of cloned, and therefore unmethylated, rDNA (Bird, 1978). These findings were consistent with studies of MspI digested calf thymus DNA. When end-labelled restriction fragments were digested to the corresponding 5'-deoxyribonucleoside monophosphates, 5-methylcytidylic acid was found in approximately 90% of the ends indicating heavy and symmetrical methylation of CpGs within the sequence CCGG in calf thymus DNA (Cedar *et al*, 1979). More recent studies have uncovered strand-specific cycling of DNA methylation at an active promoter (Métivier *et al*, 2008; Kangaspeska *et al*, 2008) although the function of these hemimethylated sites is at present unknown.

1.1.5 Methylated and unmethylated compartments of the genome

The distribution of symmetrically methylated CpGs within genomes is itself non-random. A breakthrough in the understanding of DNA methylation patterns came from studies on the sea urchin *Echinus esculentus*. Digestion with methylation-sensitive restriction enzymes separates genomic DNA into high (>15 kb) and low (0.1 – 4 kb) molecular weight fractions. Furthermore, the methylation-insensitive restriction enzyme MspI (an isoschizomer of HpaII) digested the high molecular weight fractions to fragments mainly smaller than 5 kb. These experiments, together with the fact that tritiated methionine administered to embryos specifically labelled this high molecular weight fraction, showed that the sea urchin genome contains tracts of DNA where most of the CpG dinucleotides are methylated and tracts where most of them are not (Bird *et al*, 1979). This compartmental pattern is referred to as 'mosaic methylation' and is found in a variety of other organisms such as the sea quirt *Ciona intestinalis* (Simmen *et al*, 1999; Suzuki *et al*, 2007).

The term 'global DNA methylation' is often used to describe the distribution of this modification in vertebrate genomes. In contrast to the situation in the sea urchin, digestion of chicken DNA with methylation-sensitive restriction enzymes yields no low molecular weight DNA fraction that can be visualised by ethidium bromide staining. However, end-labelling of the digested DNA followed by agarose gel autoradiography revealed that over half of the HpaII fragments were concentrated in a low molecular weight (<0.5 kb) DNA fraction (Cooper *et al*, 1983). A similar situation occurs in the mouse where the methylated compartment dominates and unmethylated CpGs are concentrated in short relatively CpG rich regions known as CpG islands, which account for only approximately 1% of the genome (Bird *et al*, 1985). A recent study used a chromatography technique to purify human DNA enriched in unmodified CpG dinucleotides. Large scale sequencing of the fractionated DNA allowed the authors to estimate the presence of between 25,000 and 30,000 CpG islands in the human genome with approximately half of the identified CpG islands overlapping a known transcription

start site (Illingworth *et al*, 2008). These experimentally determined CpG islands showed good overlap with those predicted bioinformatically by the criteria of having more than 55% G+C base composition as well as a CpG observed over expected frequency greater than 0.65 and length over 500 bp (Takai and Jones, 2002).

1.1.6 5-methylcytosine as a mutagen

Early nearest neighbour analyses revealed that CpG dinucleotides are underrepresented in the genomes of various animals (Josse *et al*, 1961; Swartz *et al*, 1962). For example, given the approximately 40% GC content of the human genome a 0.2 x 0.2 = 4% CpG frequency would be predicted. However, the observed frequency of CpG dinucleotides is actually approximately 0.8% or 5-fold less than expected. A likely explanation for this discrepancy would be inefficient repair of the T:G mismatches that arise from the spontaneous deamination of methylated cytosine to thymine. Strong support for this contention came from a study of nearest neighbour frequencies that revealed vertebrate CpG deficiencies to be matched by excesses of the dinucleotides arising from failure to repair this lesion, namely TpG and CpA (Bird, 1980). The implication of DNA methylation in CpG dinucleotide suppression is also bolstered by the observation that most of the CpG rich CpG islands in mammalian genome are usually methylation free (Bird *et al*, 1985; Illingworth *et al*, 2008). In cases where CpG rich regions are methylated, there is usually a good reason to expect their presence to be under positive selection pressure. For example, this is likely to be the case at imprinting control regions where methylation appears to regulate gene expression as will be discussed in section 1.3.5. Similar forces also seem to shape the genomes of other organisms. In the sea squirt *Ciona intestinalis*, predictions of methylation based on bioinformatic analysis of CpG composition were largely validated by bisulphite sequencing (Suzuki *et al*, 2007).

Other evidence supporting the notion that cytosine methylation causes mutations in mammalian cells includes the observation that in human DNA, restriction enzyme sites containing CpG dinucleotides are relatively much more polymorphic than ones that do

not (Barker *et al*, 1984). Furthermore, disproportionately many of the point mutations found in human genetic disorders and in the tumour suppressor gene p53 in cancer arise from cytosine to thymine transitions (Cooper and Youssoufian, 1988; Rideout *et al*, 1990). Given that cytosine methylation has been tolerated by natural selection in spite of this mutagenic load, it seems likely that DNA methylation must be required for important functions in the organisms where it is present. Section 1.3 will examine the evidence for various functions proposed for DNA methylation, but first the apparatus that lays down this modification will be considered.

1.2 DNA methylation machinery

1.2.1 Identification of DNA methyltransferases

Methylation is incorporated into genomic DNA post-synthetically (Scarano *et al*, 1965; Burdon and Adams, 1969; Sneider and Potter, 1969; Kappler, 1970). Methyltransferase activities specific for CpG dinucleotide containing DNA were originally identified in a variety of mammalian extracts (Gruenbaum *et al*, 1982; Bestor and Ingram, 1983). Purification of such an activity from mouse cells allowed peptide sequences from the protein to be obtained. Using this information the gene was cloned by screening a phage library with degenerate probes, and it was eventually named *Dnmt1*. The gene sequence revealed a C-terminal domain with homology to bacterial methyltransferases and an N-terminal domain of unknown function (Bestor *et al*, 1988).

Mouse embryonic stem cells homozygous for a null allele of *Dnmt1* retained residual DNA methyltransferase activity as well as low but stable levels of DNA methylation. Extracts from these cells were able to methylate DNA *in vitro* (albeit with much reduced efficiency) and when infected with a retrovirus, these cells still carried out *de novo* methylation of the viral LTR (long terminal repeat). Taken together these experiments indicated the presence of DNA methyltransferase activities other than Dnmt1 in mouse

cells (Lei *et al*, 1996). Searches of expressed sequence tag (EST) databases for sequences with homology to Dnmt1 led to the identification of another candidate DNA methyltransferase, Dnmt2 (Yoder and Bestor, 1998). However, convincing evidence supporting a role for Dnmt2 in DNA methylation was not forthcoming and more recent work suggests that its major role is to methylate a cytosine in an aspartic acid transfer RNA (Goll *et al*, 2006). As of yet the biological significance of this conserved function is unknown, but as the protein is required in the cytoplasm rather than the nucleus for normal differentiation of the liver and brain in zebrafish (Rai *et al*, 2007), a pivotal role as a DNA methyltransferase appears unlikely. EST database searches by different investigators also led to the identification of the *Dnmt3a* and *Dnmt3b* genes, which turned out to encode *bona fide* CpG methyltransferase activities (Okano *et al*, 1998). Finally, Dnmt3L, a protein related to the mammalian DNA methyltransferases was identified. This protein lacks conserved catalytic motifs and is involved in regulating some of the functions of DNA methyltransferases discussed in section 1.3.5.

1.2.2 Catalytic mechanism

The crystal structure of the bacterial DNA cytosine-5-methyltransferase HhaI (Klimasauskas *et al*, 1994) provided strong evidence for a previously proposed catalytic mechanism for DNA methylation (Santi *et al*, 1983). The reaction mechanism consists of a nucleophilic attack on the carbon 6 of cytosine by an electron pair from a sulphur atom of a conserved catalytic cysteine (Figure 1.2A). This step is facilitated by the protonation of the nitrogen 3 position of cytosine by a conserved glutamic acid and leads to the formation of an intermediate where the enzyme is covalently bound to its substrate and electrons are pushed across the carbon 5 position. This enamine then attacks the methyl group of the co-factor s-adenosyl-L-methionine (SAM) thus transferring the group to the carbon 5 position of cytosine (Figure 1.2B). The removal of a proton from carbon 5 leads to reformation of the double bond between carbons 5 and 6 and the release by β-elimination of the covalently bound cysteine (Figure 1.2C). Before the

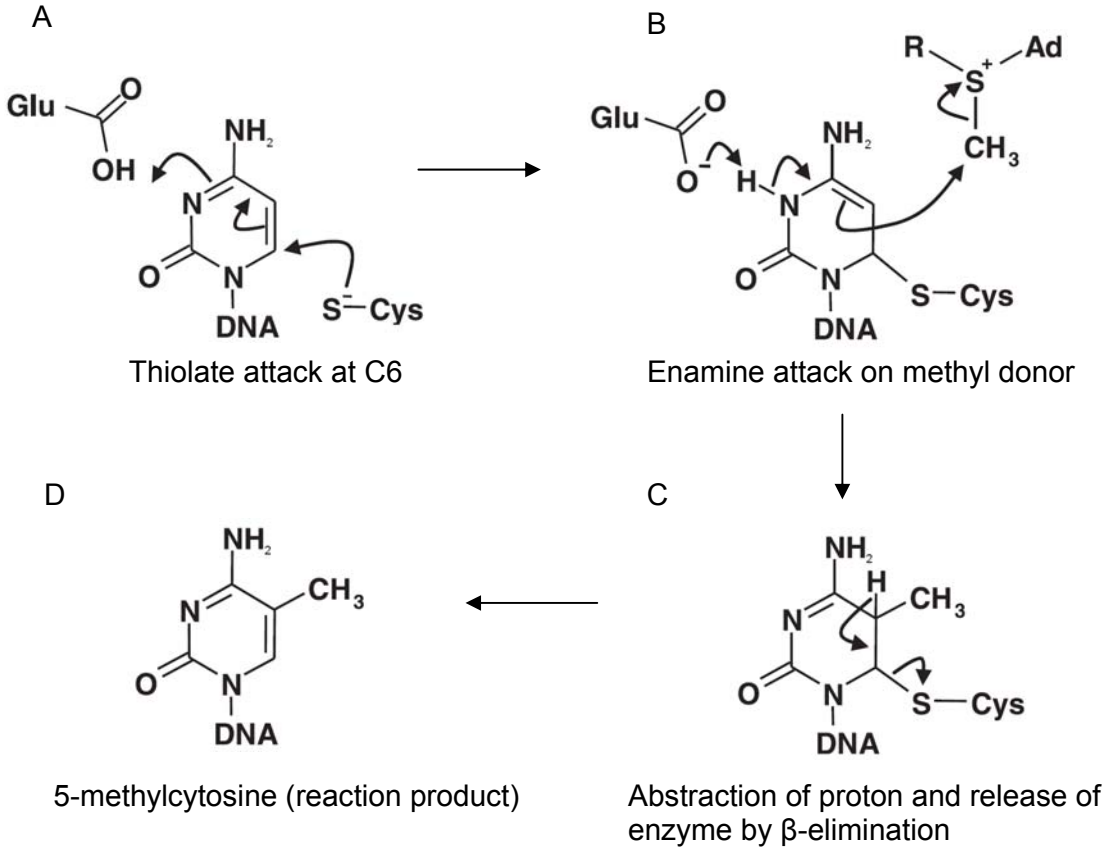

Figure 1.2 – Catalytic mechanism of cytosine methylation reaction (adapted from Jeltsch, 2006)

(A) A lone pair of electrons from a sulphur atom of a conserved catalytic cysteine in DNA methyltransferases carries out a nucleophilic attack on the carbon 6 position of cytosine. The protonation of nitrogen 3 of cytosine by an enzymatic glutamic acid promotes this reaction.
(B) The reactive 4-5 enamine formed attacks the methyl group of SAM leading to its transfer to carbon 5. This is coincident with the removal of a proton from nitrogen 3. The enzyme and substrate remain covalently bound via the sulphur atom attached to carbon 6.
(C) Abstraction of a proton from the carbon 5 position leads to reformation of the double bond between carbons 5 and 6 and release of the enzyme by a β-elimination reaction.
(D) Depiction of the free 5-methylcytosine reaction product.

crystal structure of HhaI in complex with DNA was solved, it was difficult to see how this reaction could proceed as neighbouring nucleotides lie in the way of groups which would have to react with the C5 and C6 positions of cytosine. In the structure the target base is flipped out of the double helix so that it can be positioned in the active site of the enzyme. Due to the homology between the bacterial HhaI and the mammalian methyltransferase enzymes, it is likely that the catalytic mechanism is conserved.

1.2.3 Maintenance methylation and epigenetics

The evidence discussed in section 1.1 for symmetrical CpG methylation as well as the stable presence of heavily methylated and largely unmethylated sequences in the same organism, is consistent with the idea that methylation patterns are replicated during cell division. This view was supported by experiments where unmodified or artificially methylated foreign DNA sequences were inserted into cultured mouse cells (Wigler *et al*, 1981; Stein *et al*, 1982a). Although methylation patterns were not maintained with complete fidelity, the presence or absence of CpG methylation in tracts of DNA (and not the presence of cytosine methylation in other contexts) was faithfully sustained over many cell divisions. Consistent with this imperfect copying of methylation patterns, heterogeneous methylation was found at several genomic sites in solid tumours in spite of their clonality (Silva *et al*, 1993). Furthermore, bisulphite sequencing analysis of the endogenous FMR1 promoter revealed stability in individuals in the ratio of cells containing heavily methylated versus largely unmethylated tracts. Also it was revealed that heavily methylated clones are heterogeneous with respect to which individual CpGs lack the modification (Stöger *et al*, 1997). Therefore, like foreign DNA inserted into mouse cells, endogenous sequences also seem to retain their overall methylation status through cell divisions, but with little accuracy at the level of single CpG sites.

Experiments *in vitro* revealed a strong preference of endogenous mammalian DNA methyltransferase activities as well as recombinant human DNMT1 for hemimethylated substrates (Gruenbaum *et al*, 1982; Bestor and Ingram, 1983; Pradhan, *et al*, 1999). Such

an activity would be expected to act preferentially on the hemimethylated sites generated during replication of fully methylated CpGs, and thus a mechanistic explanation for the copying of methylation patterns is provided. A number of proteins reported to interact with Dnmt1 are also likely to be involved in the copying of methylation patterns. For example, it has been proposed that the SRA (SET and RING associated) domain containing protein Np95/UHRF1 recruits the maintenance DNA methyltransferase to the hemimethylated CpG sites produced during S phase. Lines of evidence supporting this model include the interaction of recombinant Np95/UHRF1 with hemimethylated DNA *in vitro*, the observation of a physical interaction between DNMT1 and UHRF1, and the mislocalisation of Dnmt1 together with the loss of methylation seen in mouse embryonic stem cells where the *Uhrf1* gene has been ablated (Bostick *et al*, 2007; Sharif *et al*, 2007).

The fact that DNA methylation represents information aside from the genome sequence that can be inherited through cell divisions supported the earlier suggestion that this modification of the genome might be involved in events requiring cellular memory such as changes in gene expression during differentiation and X-inactivation (Holliday and Pugh, 1975; Riggs, 1975). For this reason the study of DNA methylation and related phenomena is often considered to be part of a field known as 'epigenetics' (Bird, 2002).

1.2.4 *De novo* methylation

As discussed above, DNA methylation patterns can, at least to some extent, be stably propagated through cell divisions, but these patterns must somehow be established in the first place. In mammals, global demethylation of the genome by as yet unidentified demethylase activities occurs in preimplantation embryos and primordial germ cells (Reik *et al*, 2001). Subsequent to this, *de novo* DNA methylation must occur in order to re-establish methylation patterns. Although relatively inefficient enzymes *in vitro*, recombinant Dnmt3A and Dnmt3B showed no preference for hemimethylated over unmethylated substrate DNA indicating that these might be good candidates for a role in

de novo methylation *in vivo*. Consistent with this view these enzymes are relatively highly expressed in embryonic stem cells and the early mouse embryo (Okano *et al*, 1998) where *de novo* DNA methyltransferase activity is at its highest (Jähner *et al*, 1982; Palmiter *et al*, 1982; Stewart *et al*, 1982).

Genetic studies have also strongly implicated *Dnmt3a* and *Dnmt3b* in *de novo* methylation *in vivo*. Mouse embryonic stem cells where both *Dnmt3a* and *Dnmt3b* have been ablated fail to carry out *de novo* methylation of proviral DNA following infection with a retrovirus. This is in sharp contrast to wild-type cells and even *Dnmt1*-null cells, which can still methylate proviral DNA albeit with slower kinetics. Furthermore, mouse embryos lacking *Dnmt3a* and *Dnmt3b* showed essentially identical methylation levels at C-type retroviral DNA and intra-cisternal A particle (IAP) repeats to those seen in the blastocyst. The situation is different in *Dnmt1*-deficient embryos where methylation of these sequences is significantly lower than in the blastocyst. These comparisons were taken to mean that unlike Dnmt1, the methyltransferases Dnmt3A and Dnmt3B are not required for the maintenance of existing methylation. Rather they are needed to establish embryonic methylation patterns *de novo* subsequent to the relatively hypomethylated blastocyst stage (Okano *et al*, 1999).

The model that emerged from the above lines of research proved to be too simplistic and required some modifications. Extended culture of mouse embryonic stem cells lacking both Dnmt3a and Dnmt3b gave rise to a progressive loss of CpG methylation, ultimately to levels comparable to *Dnmt1*-null cells (Chen *et al*, 2003). Taken together with the subtle defect in *de novo* methylation of proviral DNA seen in *Dnmt1* knock-out cells, it is apparent that a cut and dried distinction between Dnmt1 as a maintenance enzyme versus Dnmt3a and Dnmt3b as *de novo* methyltransferases is not appropriate. A more realistic view of maintenance methylation likely involves passive demethylation due to imperfect completion of every hemimethylated site by Dnmt1 being offset by a degree of *de novo* methylation by Dnmt3A and Dnmt3B within methylated domains.

1.2.5 Factors required for targeting DNA methylation

For a long time it was unclear how DNA methylation was targeted to certain regions of genomes whilst being excluded from others. However, recent studies have begun to throw some light upon this question. The situation is complex and the exact contribution of various factors depends on the organism being studied. Nevertheless, roles in this process are now starting to be delineated for RNA interference (RNAi), chromatin structure, DNA sequence, transcription, and a variety of proteins other than the DNA methyltransferase enzymes.

A number of factors appear to be required for the efficient function of DNA methyltransferase enzymes. For example, the lysine demethylase mLSD1 is required for proper DNA methylation in mouse embryonic stem cells, and this is reported to be because Dnmt1 is hypermethylated on a lysine residue leading to its destabilization in *Lsd1*-deficient cells (Wang *et al*, 2009). A separate interaction of DNMT1 with PCNA (proliferating cell nuclear antigen) was proposed to be important for recruiting DNA methyltransferase activity to replication forks in S phase, and thus for the efficient methylation of newly synthesised DNA (Leonhardt *et al*, 1992; Chuang *et al*, 1997). However, this interaction may only be of moderate significance as a truncated form of DNMT1 which cannot interact with PCNA is recruited to DNA with only two-fold reduced efficiency (Spada *et al*, 2007). Although roles for LSD1 and PCNA in efficient DNA methylation remain plausible, it is not clear whether these factors play any part in targeting methylation to specific genomic regions.

There is strong evidence that RNAi can direct transcriptional silencing and DNA methylation in plants. RNA-directed DNA methylation (RdDM) was first observed in tobacco plants carrying transgenic viral sequences where replication of the viral RNA induced methylation of the transgenic DNA but not the rest of the genome (Wassenegger *et al*, 1994). A similar effect was seen with a transcribed transgenic inverted repeat (Mette *et al*, 1999) thus correlating the presence of double stranded RNA with DNA

methylation at homologous sequences. The plant DNA methyltransferases responsible for the establishment of RdDM were subsequently found to be DRM and CMT3 rather than MET1, which is only involved in its maintenance (Cao *et al*, 2003; Jones *et al*, 2001). It was demonstrated that mutations in the *Arabidopsis* RNAi pathway result in failure to establish methylation at certain genomic regions (Chan *et al*, 2004) and furthermore, a putative DNA-dependent RNA polymerase involved in this process has been identified (Herr *et al*, 2005; Onodera *et al*, 2005). The small RNAs produced by DICER-LIKE3 are then incorporated into ARGONAUTE4, which is thought to recruit methyltransferase enzymes to homologous DNA sequences (Zilberman *et al*, 2004; Li *et al*, 2006; Qi *et al*, 2006). Despite the clear evidence available in plants, whether or not RNAi can direct DNA methylation in mammalian cells remains a controversial issue (Morris *et al*, 2004; Ting *et al*, 2005).

Several lines of evidence indicate that transcription might play a direct role in dictating DNA methylation patterns. Using transgenic mice it was shown that mutation of sites for the transcriptional activator Sp1 in the *Aprt* CpG island promoter leads to methylation of this region (Brandeis *et al*, 1994; Macleod *et al*, 1994). This is consistent with a model where transcription from a promoter normally keeps the region free from methylation. There is also circumstantial evidence that sequences that are transcribed through are preferentially targeted for DNA methylation. For instance, human coding regions are more likely to be methylated when they reside on the active rather than the inactive X chromosome (Hellman and Chess, 2007). Consistent with this, recent data from mouse studies suggested that transcription from nearby protein coding genes through imprinted differentially methylated regions is required for them to become methylated (Chotalia *et al*, 2009). The link between transcription and DNA methylation may prove to be extremely general. In the sea squirt methylation is often associated with gene bodies particularly at evolutionarily ancient housekeeping genes which are presumably transcribed in many tissues (Suzuki *et al*, 2007). Furthermore, coding regions also appear to be methylation targets in *Arabidopsis* (Zilberman *et al*, 2007).

Several avenues of research have implicated ATP-dependent chromatin remodelling enzymes in the targeting of DNA methylation. Genetic studies in *Arabidopsis* uncovered a requirement for DDM1 (deficient in DNA methylation 1), a SWI2/SNF2-like protein, for proper DNA methylation, especially of repetitive sequences (Vongs *et* al, 1993; Jeddeloh *et al*, 1999). Consistent with this finding, Lsh (lymphoid specific helicase), the mammalian homologue of DDM1, is also required for proper methylation in the mouse (Dennis *et al,* 2001). As well as promoting methylation of repetitive sequences, Lsh also seems to be required for the modification and silencing of certain genes. Chromatin immunoprecipitation revealed the presence of Lsh at the imprinted *Cdkn1c* locus. This gene loses methylation and is inappropriately expressed from the paternal allele in mouse embryos when *Lsh* is deleted, even though regulation of other imprinted genes is unaffected (Fan *et al*, 2005). Furthermore, Lsh is present at the *Hox* cluster and when the protein is removed there is reduced methylation and aberrant expression of some of these genes (Xi *et al*, 2007). Human cells carrying mutations in ATRX, another member of the SWI2/SNF2 family of proteins, show changes in DNA methylation patterns at rDNA arrays and several other repetitive sequences. This further suggests a role for chromatin remodelling in the targeting of DNA methylation (Gibbons *et al*, 2000).

Covalent modifications of histones can also now be regarded as important in the targeting of DNA methylation patterns. Genetic studies have shown that cytosine methylation in *Neurospora crassa* is dependent on the trimethylation of histone H3 at lysine K9 by DIM-5 (Tamaru and Selker, 2001; Tamaru *et al*, 2003). Furthermore, the *Neurospora* homolog of HP1, a protein that binds to the N-terminal tail of histone H3 only when it is methylated at lysine K9, is also required for DNA methylation (Freitag *et al*, 2004). The direct physical interaction observed between this protein and the *Neurospora crassa* DNA methyltransferase enzyme suggested a model where HP1 recruits this enzyme to chromatin containing histone H3 methylated at lysine K9 (Honda and Selker, 2008). Mutations in *KRYPTONITE* from *Arabidopsis*, which encodes an H3 lysine K9 methyltransferase, also cause a reduction in cytosine methylation levels arguing for a more general role for this histone mark in promoting DNA methylation

(Jackson *et al*, 2002). Furthermore, in mouse cells deficient for Suv39h1 and Suv39h2, two mammalian histone H3 lysine K9 methyltransferase enzymes, DNA methylation is reduced at pericentric heterochromatin. An interaction between Dnmt3b and HP1α together with the mis-localisation of Dnmt3b in these mutant cells suggests a similar mechanism is at work to the one described in *Neurospora* (Lehnertz *et al*, 2003).

A case can now also be made for a role of histone H3 lysine K4 methylation in controlling which genomic regions are subject to DNA methylation. The N-terminal cysteine-rich domain of Dnmt3L was recently shown to associate with the N-terminal tail of histone H3 but only when it is not methylated at lysine K4 (Ooi *et al*, 2007). As Dnmt3L is a non-catalytic regulator of *de novo* DNA methyltransferases that is required for methylation of maternally imprinted loci and of retrotransposons in male germ cells (Bourc'his, *et al*, 2001; Bourc'his and Bestor, 2004), this suggests that the absence of methylation at lysine K4 of histone H3 might permit the recruitment of DNA methyltransferase activity to specific parts of the genome.

Finally, the idea that specific DNA sequences (beyond the simple requirement for CpG dinucleotides) might govern methylation patterns by being better or worse substrates for methyltransferase enzymes was recently brought into focus (Jia *et al*, 2007). The crystal structural of a tetrameric complex of Dnmt3a and Dnmt3L revealed that the two active sites are separated by 40 Å approximately corresponding to one helical turn of DNA. This suggested that DNA methylation might be regulated by a preference of Dnmt3a for sequences with CpGs separated by approximately 10 base pairs or one helical turn. Supporting this idea, a periodicity of 8-10 base pairs was found in the methylation of stretches of DNA incubated with recombinant Dnmt3a *in vitro* and a similar periodicity was found in the CpG dinucleotides present in differentially methylated regions of maternally imprinted genes, but not randomly selected CpG islands. Given that Dnmt3a and Dnmt3L are required in germ cells for the establishment of imprints (Kaneda *et al*, 2004; Bourc'his *et al*, 2001), an argument can be made that sequence preferences of methyltransferases contribute to the targeting of DNA methylation.

1.2.6 Aberrant DNA methylation

As will be discussed in section 1.3, heavy methylation of promoter DNA is incompatible with transcription of the associated gene. Although DNA from tumour samples tends to be globally hypomethylated (Feinberg and Vogelstein, 1983; Gama-Sosa *et al*, 1983), promoters of tumour suppressor genes are frequently observed to be silenced and methylated in cancer giving rise to the suggestion that DNA methylation plays a role in carcinogenesis (Jones and Baylin, 2002). What remains unclear, however, is whether DNA methylation is a cause or consequence of such transcriptional silencing. Even if it is assumed to be causal, it is unknown whether methylation of tumour suppressor genes represents random aberrant events driving cancer development, or if it is dictated by genetically induced changes in oncogenes lying upstream in tumorigenesis.

Evidence exists supporting each of the viewpoints described above. When different mouse cell lines are grown in culture many of the CpG islands of a similar set of dispensable genes acquire methylation (Antequera *et al*, 1990). This argues that methylation is either targeted to promoters that are already repressed or else it occurs randomly and is only selected against when it coincides with and represses the promoter of a gene required for proliferation. Consistent with the first interpretation most of the genes found to be methylated in colon cancer are also found to be transcriptionally inactive in the normal colon (Keshet *et al*, 2006). However, the second interpretation fits in with the fact that demethylating agents such as 5-azadeoxycytidine can often partially reactivate silenced genes in cancer (Jones and Baylin, 2002). Experiments favouring a model where activated oncogenes direct DNA methylation and silencing to promoters of specific tumour suppressors have been carried out by researchers studying Ras. In this case an elaborate pathway to specifically methylate and suppress certain growth inhibitory genes seems to be hijacked by Ras in the transformation process (Gazin *et al*, 2007). In the light of these data, it seems most plausible that DNA methylation can be part of the oncogenic process both by being co-opted by activated oncogenes to silence

specific promoters, and by providing a source of random yet heritable changes in gene expression profiles.

1.3 Functions of DNA methylation

1.3.1 DNA methylation is essential for proper development in mammals

Mice deficient for the maintenance methyltransferase enzyme Dnmt1 are not viable (Li *et al*, 1992) and when they are homozygous for a null allele of *Dnmt1* development is arrested prior to the 8-somite stage (Lei *et al*, 1996). In addition to its enzymatic activity Dnmt1/DNMT1 associates with a variety of other proteins such as histone deacetylases and can function as a transcriptional repressor in some assays (Fuks *et al*, 2000; Robertson *et al*, 2000). However, it appears that the primary cause of the phenotype observed in the absence of Dnmt1 is the reduction in DNA methylation as embryonic stem cells expressing Dnmt1 with a conservative point mutation that abolishes its enzymatic activity show similar defects to those not expressing the protein at all (Damelin and Bestor, 2007). The *de novo* DNA methyltransferase enzymes are also critical for normal mammalian development. *Dnmt3a*-deficient mice appear normal at birth but soon become runted and usually die within four weeks. *Dnmt3b* mutants die in gestation showing developmental abnormalities including growth retardation and rostral neural tube defects (Okano *et al*, 1999). *DNMT3B* is also important for normal human development with mutations in this gene giving rise to ICF (immunodeficiency, centromere instability and facial anomalies) syndrome (Hansen *et al*, 1999). As many ICF mutations affect the catalytic domain of DNMT3B (Xu *et al*, 1999) and the satellite DNA of ICF patients is highly demethylated (Jeanpierre *et al*, 1993), it appears that the developmental defects here are likely to be down to loss of methyltransferase activity rather than some other function of the protein.

Although it is clear that DNA methylation is required for normal development, to date many of the molecular details of how a lack of DNA methylation leads to the defects described above remain to be elucidated. Many lines of evidence point to a potential role for DNA methylation in transcriptional repression and these will be considered later. First, however, a potential role for DNA methylation in the maintenance of genomic stability will be discussed.

1.3.2 Genome stability

Early observations of human cells administered with the demethylating agent 5-azadexoycytidine revealed increased chromosome associations between satellites, secondary constrictions and telomeric regions (Viegas-Péquignot and Dutrillaux, 1976). Similar abnormal associations were seen between methylation deficient centromeric regions in mitogen stimulated cells from ICF patients (Jeanpierre *et al*, 1993; Xu *et al*, 1999). Together these results suggested that DNA methylation might function in the maintenance of chromosome integrity. However, this conclusion should be treated with caution as any effects observed in experiments using 5-azadeoxycytidine could be due to its ability to crosslink proteins to DNA (Jüttermann *et al*, 1994) and the chromosome rearrangements observed in ICF are not seen in cells taken directly from patients (Xu *et al*, 1999).

Genetic studies in mice have also asked whether DNA methylation might be involved in maintaining genomic stability. Embryonic stem cells null for *Dnmt1* were reported to exhibit increased mutation rates and the types of mutations observed suggest a role for methylation in the suppression of mitotic recombination as well as faithful chromosome segregation during mitosis (Chen *et al*, 1998). Also supporting a role for methylation in the maintenance of genome stability, mice hypomorphic for *Dnmt1* develop aggressive T cell lymphomas aged 4-8 months. These tumours often displayed trisomy of chromosome 15 indicating that demethylation could lead to carcinogenesis by promoting chromosome instability (Gaudet *et al*, 2003). However, once again these conclusions

must be treated with a degree of scepticism. An independent study found no evidence of an elevated mutation rate in *Dnmt1*-null embryonic stem cells (Chan *et al*, 2001) and $Apc^{Min/+}$ mice develop fewer polyps, perhaps due to a decreased rate of loss of heterozygousity, when they are hypomorphic for *Dnmt1* (Laird *et al*, 1995).

Less disputed evidence that DNA methylation is important for genomic stability comes from studies of the fungus *Neurospora crassa*. In this organism complete loss of methylation due to mutation of the DNA methyltransferase gene dim-2 or partial loss of methylation induced by S-adenosylmethionine starvation leads to aneuploidy (Foss *et al*, 1993). However, it is difficult to be confident that this is a direct consequence of DNA methylation defects rather than being a secondary effect of changes in gene expression due to loss of methylation. The same caution would have to be applied to mammalian systems even if the data supporting a role for DNA methylation in the maintenance of genome stability discussed here is taken at face value.

1.3.3 Transcriptional repression

A lot of early work suggested that DNA methylation might play a role in transcriptional repression. An inverse correlation was noted between methylation and expression levels of integrated viral genes in adenovirus-transformed cell lines (Sutter and Doerfler, 1980; Vardimon *et al*, 1980) and in support of a causal relationship cloned adenovirus E2a region microinjected into *Xenopus* oocytes was repressed specifically when methylated (Vardimon *et al*, 1982). Consistent with this result, artificially methylated transgenes inserted into mouse cells failed to be expressed compared to unmethylated controls (Stein *et al*, 1982b) and promoter CpG methylation also inhibited transcription of cloned genes in transfected cells and in transcription assays *in vitro* (Ben-Hattar and Jiricny, 1988; Watt and Molloy, 1988; Iguchi-Ariga and Schaffner, 1989). Treatment of certain cell types with 5-azadexoycytidine led to the reactivation of some genes (Jones, 1985) implicating methylation in the repression of endogenous promoters, however, there are caveats associated with experiments using this drug that are mentioned above. Also it

must be kept in mind that cell lines often methylate promoters that are normally methylation free in tissues (Antequera *et al*, 1989) and so such experiments using cultured cells may reveal a role for methylation in gene silencing that is not present under physiological conditions.

Later work firmly implicated methylation-mediated gene silencing in X-inactivation and imprinting as will be discussed in sections 1.3.4 and 1.3.5. However, compelling evidence for a general role for DNA methylation in gene repression under physiological conditions is still lacking. A recently described example which may prove to be an exception to this is the repression of the transcription factor gene *Elf5* in embryonic stem cells (Ng *et al*, 2008). *Elf5* is needed for the proliferative capacity of trophoblast cells and is aberrantly expressed in methylation deficient embryonic stem cells, where its promoter becomes demethylated. This results in aberrant differentiation of embryonic stem cells into trophoblast giant cells implying that DNA methylation controls development by regulating gene expression. Work on the endogenous *Oct4* gene that is silenced and methylated during development, however, showed that DNA methylation occurs only after silencing and serves merely to stabilize the repressed state (Feldman *et al*, 2006). Whether this paucity of examples of genes being developmentally regulated by DNA methylation reflects the true biological role of methylation or the technical limitations of current methodologies is unclear.

DNA methylation-mediated transcriptional repression has been shown to regulate parasitic repetitive DNA elements. Some relatively benign parasitic sequences transpose at high frequencies when cloned copies are transfected into human cells (Moran *et al*, 1996) and IAP retroviral transcripts are upregulated over 50-fold in *Dnmt1*-deficient mouse embryonic stem cells (Walsh *et al*, 1998). Furthermore, tumours from mice hypomorphic for *Dnmt1* are frequently found to express oncogenic Notch1 due to insertional mutagenesis by an IAP retrovirus (Howard *et al*, 2008). Studies in *Arabidopsis* also support the notion that DNA methylation functions to protect genomes from parasitic elements as mutations in this organism that cause decreased methylation

also lead to mobilization of transposons (Miura *et al*, 2001). Nevertheless the conclusion that a primary function of DNA methylation is to protect genomes from mutagenic transposition by parasitic elements is controversial. Some organisms such as the sea squirt *Ciona intestinalis* appear to target methylation to gene bodies rather than to transposable elements (Simmen *et al*, 1999; Suzuki *et al*, 2007). Also it has been argued that the hypomethylation of transposable elements observed in oocytes, testis and early mouse embryos, precisely at the time when transposition could be most damaging, makes it unlikely that the main purpose of DNA methylation is to suppress mobilization of these elements. An alternative but not mutually exclusive possibility is that transcription of parasitic elements represents 'transcriptional noise' which could interfere with the precise gene expression patterns required in various differentiated cell types (Bird, 1997). Corroborating evidence for this hypothesis is, however, lacking at present.

1.3.4 X-inactivation

Mammalian females carry two copies of the X chromosome whereas males only have one. In order to equalize the expression levels of X-linked genes between the two genders females silence one copy of the X chromosome in each cell. Early in the development of placental mammals each cell of a female embryo randomly chooses one copy of the X chromosome for inactivation and then this decision is stably propagated in all of the daughters of that cell. This model was proposed partly on the basis of the mosaic phenotypes of female mice heterozygous for a variety of sex-linked mutations (Lyon, 1961) and has proved to be largely correct, only breaking down in the very early embryo, where the paternal X chromosome is subject to imprinted X inactivation (Mak *et al*, 2004; Okamoto *et al*, 2004). This state persists throughout development in the extraembryonic lineages, but in the embryo proper the silenced paternal X chromosome is reactivated before the onset of random X inactivation. Given that X inactivation involves a heritable epigenetic distinction between identical sequences present in the same cell, the involvement of DNA methylation in this process now seems like an

obvious hypothesis. Remarkably, however, this was predicted before CpG methylation was known to occur predominantly symmetrically and before any eukaryotic maintenance methyltransferase activities had been identified (Riggs, 1975).

Support for this idea came from experiments using 5-azadexoycytidine. By using this drug to induce genomic demethylation in interspecies somatic cell hybrids it was possible to reactivate genes residing on the inactive X chromosome (Mohandas *et al*, 1981) and this occurred concomitantly with the demethylation of the promoter of a reactivated gene (Hansen and Gartler, 1990). Evidence also emerged of differential methylation between the active and inactive X chromosomes. Digestion of human metaphase chromosomes with the methylation sensitive restriction enzyme HhaI followed by end-labelling with biotinylated dUTP and immunofluorescence revealed higher methylation levels on the active than the inactive X chromosome (Viegas-Péquignot *et al*, 1988) and this observation was subsequently verified by microarray analysis of fragmented genomic DNA fractionated with an antibody against methylated cytosine (Weber *et al*, 2005). These results are somewhat surprising given the association between methylation and transcriptional repression. However, CpG islands on the inactive X chromosome do specifically acquire methylation consistent with a model where DNA methylation represses transcription from promoters on the inactive X (Wolf *et al*, 1984; Lock *et al*, 1986). Further supporting the link between methylation and repression is the fact that the CpG islands of genes which escape X-inactivation are infrequently methylated compared to those which are silenced (Carrel and Willard, 2005; Illingworth *et al*, 2008). Genetic studies also support a role for DNA methylation in regulating mammalian dosage compensation. The expression pattern of X-linked lacZ transgenes in *Dnmt1*-deficient embryos revealed defective random X chromosome inactivation where DNA methylation is lacking (Sado *et al*, 2000).

One feature of the role of DNA methylation in the regulation of X inactivation is that it appears to be required for the maintenance rather than the establishment of gene silencing. This is demonstrated by the fact that the X-linked mouse *Hprt* gene promoter

only becomes methylated after inactivation has taken place (Lock *et al*, 1987). This facet is also evident in the methylation mediated silencing of the non-coding Xist RNA promoter, which is mis-regulated in methylation-deficient differentiating *Dnmt1*-null cells but not in the corresponding embryonic stem cells (Beard *et al*, 1995). The Xist RNA is expressed exclusively from the inactive X chromosome, which it coats and is required for the establishment of silencing of in *cis* (Penny *et al*, 1996). Aberrant expression of the Xist RNA from the male active X chromosome in methylation-deficient cells results in down regulation of many X-linked genes and may contribute to the apoptotic phenotype of differentiating *Dnmt1*-null embryonic stem cells (Panning and Jaenisch, 1996).

1.3.5 Imprinting

Imprinted genes are expressed preferentially from either the maternal or paternal allele with conservative estimates predicting around one hundred mammalian genes likely to be regulated in this way (Reik and Walter, 2001). The first, and best studied, genes demonstrated to show imprinted expression profiles were the mouse *Igf-2* and *H19* loci, which are transcribed exclusively from the paternal and maternal alleles respectively (DeChiara *et al*, 1991; Bartolomei *et al*, 1991). Maternal inheritance of a deletion in the *H19* gene leads to biallelic expression of *Igf-2* (Leighton *et al*, 1995) implying that imprinting of these two loci, which lie only ninety kilobases apart (Zemel *et al*, 1992), is somehow connected. *Igf-2* imprinting is independent of the presence of the H19 RNA (Jones *et al*, 1998) suggesting that a common regulatory mechanism exists for these two genes rather than the repression of one simply being downstream of the expression of the other.

As was the case in X-inactivation, DNA methylation represented a strong candidate for a role in the regulation of imprinting, due to a need for heritable discrimination of two copies of the same sequence in one cell. Early evidence supporting this notion was the observation that foreign DNA inserted into particular mouse genomic regions would

show parent of origin-specific differences in DNA methylation levels (Sapienza *et al*, 1987; Reik *et al*, 1987). This was backed up by analysis of the endogenous *H19* locus, which is specifically methylated on the paternal allele (Ferguson-Smith *et al*, 1993; Bartolomei *et al*, 1993). Ultimately a causal role for DNA methylation in imprinting was established by genetic studies in mice.

Mouse embryos null for *Dnmt1* show aberrant biallelic expression of the *H19* gene and silencing of the normally expressed paternal *Igf-2* allele (Li *et al*, 1993). The finding that a deficiency in methylation can lead to erroneous gene repression may seem somewhat paradoxical given the implication of DNA methylation in transcriptional silencing. However, this was resolved by the discovery of a distant enhancer that normally only stimulates expression of the paternal *Igf-2* allele. Such specificity arises because the maternal allele is shielded from this enhancer by the binding of the insulator protein CTCF to the unmethylated allele of the closely linked *H19* differentially methylated region (Bell and Felsenfeld, 2000; Hark *et al*, 2000). Ablation of *Dnmt3L* in the mouse reinforced the evidence for a role for DNA methylation in imprinting. Heterozygous progeny of female mice null for *Dnmt3L* die mid-gestation and bisulphite sequencing revealed specific demethylation of differentially methylated imprinting control regions in maternal oocytes and embryos. This correlated with deregulated expression of imprinted genes and is the likely cause of the maternal effect lethality (Bourc'his, *et al*, 2001). Conditional knock-out of *Dnmt3a* in mouse germ cells revealed a similar role for this gene in the establishment of imprints. Progeny of females lacking *Dnmt3a* in their germ cells die *in utero* with a lack methylation and biallelic expression of all maternally imprinted genes tested (Kaneda *et al*, 2004). Finally, except for the modification of imprinted regions, rescue of *Dnmt1*-null embryonic stem cells with a *Dnmt1* cDNA largely restored genomic cytosine methylation. These cells fail to properly regulate imprinted genes thus firmly implicating a need for DNA methylation rather than some other function of the methyltransferase enzymes in this process (Tucker *et al*, 1996).

An initial reason for supposing that some genes might show parent of origin-dependent expression profiles was the early death of mouse embryos derived from zygotes reconstituted with two maternal genomes. Defective imprinting did indeed prove to be the only major barrier to the development of parthenogenetic mice. This is demonstrated by the fertile adult mice that can be generated in this fashion if one of the mothers bears deletions in certain imprinted regions such that the expression profile generated by her contribution to the progeny's genome more closely resembles a paternal one (Kono *et al*, 2004; Kawahara *et al*, 2007). Given that imprinting is disrupted in DNA methylation-deficient mice, and given that the severe phenotype seen in parthenogenetic embryos is caused by defective imprinting alone, it seems likely that mis-regulation of imprinted genes may contribute strongly to the lethal phenotype of *Dnmt1*-deficient mice.

1.4 Mediators of the methylation signal

1.4.1 Identification of methyl-CpG binding proteins

A major avenue of research in the DNA methylation field has focussed upon the identification and characterization of proteins that bind specifically to methylated DNA. *A priori* it is possible to imagine that DNA methylation mediates its effects by directly blocking the binding of proteins to DNA as is the case with CGBP, CTCF and c-Myc (Voo *et al*, 2000; Hark *et al*, 2000; Prendergast and Ziff, 1991). However, another theoretically possible mechanism is that proteins exist which can bind specifically to methylated DNA and mediate the effects of this modification by directly occluding other factors or by recruiting activities to somehow modify the local chromatin structure. The slow kinetics of repression of a transfected methylated gene suggested the effect was not simply due to blocking the binding of a transcriptional activator (Buschhausen *et al*, 1985). Furthermore, methylated CpGs in mammalian nuclei were found to be specifically protected from restriction digest implying tight association with proteins specific for these sites (Antequera *et al*, 1989).

The first activity observed to be able to bind specifically to DNA containing methylated CpG sites was found in nuclear extracts of a variety of mammalian tissues and cell lines by electrophoretic mobility shift assay (EMSA). This activity was designated MeCP1 (Meehan *et al*, 1989) and was indistinguishable from the mediator of methylation-dependent transcriptional repression in transfected cells and in nuclear extracts *in vitro* (Boyes and Bird, 1991). However, the molecular identities of the proteins comprising MeCP1 initially remained unknown. A second methyl-CpG specific binding activity distinct from MeCP1 was uncovered in rat brain nuclear extracts using a Southwestern assay. Purification of this activity allowed a partial amino acid sequence to be obtained and this led to the cloning of the gene encoding MeCP2 (Lewis *et al*, 1992). A subsequent deletion analysis of recombinant MeCP2 identified a minimal region required for specific recognition of methylated DNA and this was termed the methyl-CpG binding domain or the MBD (Nan *et al*, 1993).

Having defined the methyl-CpG binding domain, searches of EST databases led to the identification of a family of MBD containing proteins. The human *MBD1* gene was the first to be cloned (Cross *et al*, 1997) and this was quickly followed by the *Mbd1*, *Mbd2*, *Mbd3* and *Mbd4* genes from mouse (Hendrich and Bird, 1998). These reports also revealed that like MeCP2, the newly identified Mbd1, Mbd2, and Mbd4 proteins could all bind specifically to methylated DNA *in vitro*, with only Mbd3 not showing this property. It was later shown that the *Xenopus laevis* homolog of Mbd3 is able to bind specifically to methyl-CpG containing probes (Wade *et al*, 1999) although the biological significance of this evolutionary difference remains unclear.

1.4.2 Structure of the MBD

Nuclear magnetic resonance (NMR) studies of the isolated MBD domains of MeCP2 and MBD1 revealed a wedge shaped α/β-sandwich structure comprising a four strand anti-parallel β-sheet attached to a short α-helix (Wakefield *et al*, 1999; Ohki *et al*, 1999). The structure of the MBD from MBD1 was later solved in complex with methylated

DNA by NMR (Ohki *et al*, 2001). Upon DNA binding a nine amino acid loop between the inner strands of the β-sheet becomes structured and together with the inner β-strands and the loop connecting to the α-helix this forms the majority of the interface with the major groove of the DNA. It was further suggested on the basis of this structure that the methyl groups on cytosine are recognised by contacts with specific hydrophobic amino acids. However, a more recent X-ray crystal structure of the MBD from MeCP2 in complex with methylated DNA cast doubt upon some of these conclusions (Ho *et al*, 2008). While the overall fold of the domain is not challenged, hydrophobic contacts between the MBD and the methyl groups on cytosine were not observed. Rather it was suggested that water molecules, which it is known can be positioned in the major groove by 5-methylcytosine (Mayer-Jung *et al*, 1998), make specific contacts with conserved hydrophilic residues of the MBD.

1.4.3 MBD1

MBD1 will be discussed in detail in section 1.5.

1.4.4 MeCP2

As described above, MeCP2 was originally identified based on its ability to bind to methyl-CpG containing DNA *in vitro*. Evidence also emerged that MeCP2 functions in this fashion *in vivo*. In mouse cells the protein localizes to pericentric heterochromatin where over 40% of the genomic 5-methylcytosine resides, and this localization is lost either when the MBD is abrogated or cells deficient in DNA methylation are used (Nan *et al*, 1996). Furthermore, MeCP2 was found by chromatin immunoprecipitation to be present *in vivo* exclusively at the maternal methylated allele of the imprinted *U2af1-rs1* gene (Gregory *et al*, 2001).

Early studies on MeCP2 suggested that it might function as a methylation-dependent transcriptional repressor by recruiting histone deacetylase activities. MeCP2 was able to

repress methylated promoters *in vitro* and in transfection based reporter assays *in vivo*. A transcriptional repression domain was mapped to a central region of the protein using this second kind of assay (Nan *et al*, 1997) and consistent with the fact that repression was sensitive to the deacetylase inhibitor trichostatin A (TSA) this region was found to interact with mSin3A (Nan *et al*, 1998), which is a component of a histone deacetylase-containing co-repressor complex (Zhang *et al*, 1997).

A further clue to the function of *MeCP2* came from the finding that this gene is mutated in Rett syndrome, a severe autism-spectrum neurological disorder (Amir *et al*, 1999). Targeted mutations of *Mecp2* in mice gave rise to Rett syndrome-like phenotypes (Chen *et al*, 2001; Guy *et al*, 2001) and provided a tool with which to study *Mecp2* function in a biologically relevant setting. Given that MeCP2 can function as a repressor, it might have been anticipated that the phenotypes seen in its absence were due to the aberrant expression of genes normally silenced in the brain. One candidate gene, *Bdnf*, was found to be slightly upregulated in resting *Mecp2*-deficient neurons in culture (Chen *et al*, 2003; Martinowich *et al*, 2003). However, the significance of this finding is unclear as *Bdnf* is actually down regulated in the brains of *Mecp2* mutant mice and transgenic over-expression of *Bdnf* can ameliorate some of the Rett-like symptoms seen in these animals (Chang *et al*, 2006). Evidence has been presented to suggest that Mecp2 might repress various other promoters, but it is seldom clear what contribution, if any, mis-expression of these genes makes to the Rett-like phenotype (Horike *et al*, 2005; Nuber *et al*, 2005; Kriaucionis *et al*, 2006). Moreover, extensive microarray analysis of *Mecp2*-deficient mouse brains revealed only subtle changes in the expression levels of a handful of genes (Tudor *et al*, 2002). A more recent analysis of hypothalami from wild-type, *Mecp2*-null and *Mecp2* over-expressing mice revealed that the protein seems to up regulate more genes than it down regulates leading to the notion that Mecp2 may function as a transcriptional activator (Chahrour *et al*, 2008).

As MeCP2 functions as a repressor in various assays it is perhaps surprising that over-expression of genes normally silenced in the brain does not readily account for the

defects observed in Rett syndrome and *Mecp2*-null mice. It remains possible that the brain is particularly sensitive to subtle changes in gene expression and that the small alterations documented do in fact underlie the Rett-like phenotype. Alternatively, gene expression changes that account for the defects may have been missed if they only occur in a subset of cells present in the whole brains or hypothalami analysed. Nevertheless, the idea that MeCP2 functions as something other than a transcriptional regulator in the maintenance of normal neurological function must now be entertained. For example, an RNA-mediated interaction between MeCP2 and the splicing factor YB1 was uncovered and this led to the observation of splicing defects in *Mecp2* deficient mice. However, these effects are very subtle and it is not clear what contribution, if any, they make to the phenotype of these animals (Young *et al*, 2005). Other studies of MeCP2 have also largely failed to reveal why this protein is required for normal neuronal function. A variety of MeCP2 binding partners aside from mSin3A and YB1 have been reported including the histone deacetylase containing N-CoR complex as well as the chromatin remodelling enzymes Brahma and ATRX (Kokura *et al*, 2001; Harikrishnan *et al*, 2005; Nan *et al*, 2007). It is not clear in what fashion any of these interactions contribute to the normal function of MeCP2 and some of them are hotly contested (Hu *et al*, 2006; Harikrishnan *et al*, 2006). As MeCP2 is largely monomeric and has no stoichiometric binding partners in rat brain nuclear extracts (Klose and Bird, 2004) using biochemistry to search for the function of this protein is likely to be a great challenge.

1.4.5 MBD2

Subsequent to the cloning of the mouse *Mbd2* gene, the protein encoded by the human homolog was shown to be a stable component of the histone deacetylase-containing nucleosome remodelling (NuRD) complex, which corresponded to the previously identified MeCP1 activity (Zhang *et al*, 1999). Transfection based reporter assays showed that MBD2 could function as a histone deacetylase and chromatin remodelling-dependent transcriptional repressor (Ng *et al,* 1999; Feng and Zhang, 2001). Consistent with this, *in vitro* analysis showed that MBD2 promoted the binding of the NuRD

complex to nucleosomes containing methylated DNA, which were subsequently preferentially remodelled and deacetylated. These initial works had also suggested that MBD2 and MBD3 co-existed in the same NuRD complex, but more recent findings demonstrated the presence of distinct MBD2- and MBD3-containing NuRD complexes (Le Guezennec *et al*, 2006).

The fact that MBD2 is a critical component of the MeCP1 activity has further been confirmed by analysis of *Mbd2*-deficient mice where this activity is absent. The generation of these mice, which are both viable and fertile, has facilitated the study of the normal function of Mbd2. The first phenotype to be reported in these animals was defective maternal behaviour (Hendrich *et al*, 2001) but how the absence of Mbd2 leads to this deficit remains unclear. It has also been observed that *Mbd2*-mutant mice show abnormal differentiation of helper T cells and this has been attributed to defective silencing of the *Il-4* gene (Hutchins *et al*, 2002). Mbd2 also appears to be required in the colon to repress genes that are normally only expressed in the exocrine pancreas and duodenum (Berger *et al*, 2007) and another promoter identified as being directly repressed by Mbd2 is that of the *Xist* gene (Barr *et al*, 2007). In all of these cases there is evidence that Mbd2 represses transcription by recruiting histone deacetylase activity to promoters where DNA methylation is present. Therefore, in conjunction with the weight of evidence associating MBD2 with the NuRD complex, there is a strong case that Mbd2 functions primarily to repress transcription from promoters which contain methylated sequences. It will be of interest to determine the complete set of promoters bound and regulated by Mbd2 through development.

1.4.6 MBD4

Mbd4 was initially shown to interact with methylated rather than unmethylated DNA *in vitro* and over-expressed Mbd4 was localized to pericentric heterochromatin as would be expected for a methyl-CpG binding protein (Hendrich and Bird, 1998). A later study found that the MBD of this protein bound most efficiently to m^5CpG:TpG mismatches,

which are the result of 5-methylcytosine deamination. Furthermore, the C-terminal glycosylase domain of MBD4 could efficiently remove the thymine base from such mismatches (Hendrich *et al*, 1999). This suggested that MBD4 may function as a DNA repair protein to protect the genome from the damage induced by the presence of DNA methylation. This hypothesis was confirmed by the three-fold increase in cytosine to thymine transitions in *Mbd4*-null mice and the consequent acceleration in tumour formation on an $Apc^{Min/+}$ background (Millar *et al*, 2002). MBD4 has also been reported to be able to repress transcription and to bind to Sin3A and HDAC1 as well as being present at the methylated promoters of the p16 and hMLH1 genes (Kondo *et al*, 2005). However, evidence of gene regulation defects when MBD4 is removed is lacking.

1.4.7 Kaiso and Kaiso-like proteins

Aside from the MBD family there is a second group of proteins that specifically bind to methylated CpGs. Kaiso, the founder member of this group which has an N-terminal POZ domain and C-terminal zinc fingers, was originally purified from K652 cells as an activity that could bind specifically to methylated DNA (Prokhortchouk *et al*, 2001). As well as having affinity for methylated DNA, the zinc fingers of Kaiso can bind to a consensus sequence that lacks CpG dinucleotides (Daniel *et al*, 2002). However, in reporter assays, transfected Kaiso can still repress transcription in a methylation-dependent manner (Prokhortchouk *et al*, 2001). There is also evidence that Kaiso can function to silence endogenous genes by recruiting histone deacetylase activity to methylated promoters (Yoon *et al*, 2003). Purification of the N-CoR complex from HeLa cells identified Kaiso as an associated protein and siRNA-mediated depletion of Kaiso resulted in loss of N-CoR from certain methylated promoters concomitant with aberrant gene expression. In *Xenopus*, Kaiso deficiency allows zygotic transcription to commence before the mid blastula transition with subsequent phenotypes including developmental arrest and apoptosis (Ruzov *et al*, 2004). However, *Kaiso*-null mice show no developmental abnormalities or changes in gene expression profiles (Prokhortchouk

et al, 2006). Therefore, as with MeCP2, the status of Kaiso as a transcriptional repressor in mammals is not as certain as it once seemed.

Two more proteins, ZBTB4 and ZBTB38, were identified on the basis of homology with Kaiso (Filion *et al*, 2006). These proteins were also shown to be able to bind to methylated CpGs via their zinc finger domains and to repress transcription in reporter assays. The localization of these proteins to pericentric heterochromatin is also consistent with their binding to methyl-CpG and, furthermore, ZBTB4 and ZBTB38 as well as Kaiso were shown by chromatin immunoprecipitation to be present exclusively at the methylated paternal allele of the *H19/Igf2* differentially methylated region. However, no causal role in imprinting was demonstrated. A full understanding of the functions of these Kaiso-like proteins will require the generation and analysis of mice lacking these factors. A schematic of all known MBD-containing and Kaiso-like proteins is shown below (Figure 1.3).

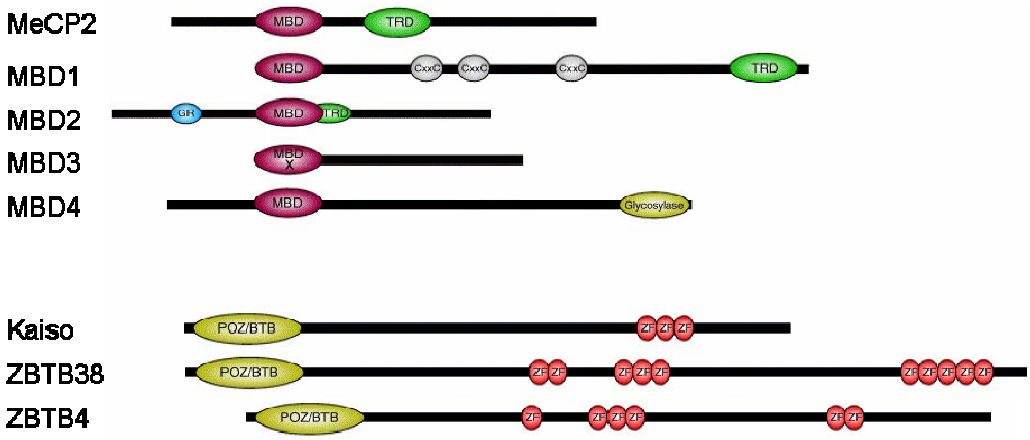

Figure 1.3 – Domain structure of methyl-CpG binding proteins

MeCP2, MBD1 and MBD2 all contain a TRD. MBD1 additionally has zinc finger domains and MBD2 has an N-terminal glycine-arginine repeat region. The MBD of MBD3 contains a substitution such that it does not bind methylated DNA. MBD4 contains a C-terminal glycosylase domain. Kaiso and the Kaiso-like proteins, ZBTB38 and ZBTB4, have a conserved POZ/BTB domain as well as three central zinc fingers that mediate interaction with methylated DNA.

1.5 The methyl-CpG binding domain protein MBD1

1.5.1 Identification of MBD1

MBD1 was first identified in searches of EST databases for proteins bearing similarity to the MBD of MeCP2. The first human isoform to be uncovered was designated PCM1, and in addition to its N-terminal MBD this protein contains two central zinc finger domains with homology to those found in the DNA methyltransferase DNMT1 and HRX proteins. An antibody raised against MBD1 was reported to 'supershift' the MeCP1 activity in an EMSA and so it was initially concluded that MBD1 was a component of the MeCP1 complex (Cross *et al*, 1997). However, later studies using more specific antibodies demonstrated that this conclusion was erroneous and may have arisen due to cross-reactivity of the original antibody with some other component of the MeCP1 complex (Ng *et al*, 1999; Ng *et al*, 2000). MBD1 was also independently identified in a yeast one-hybrid screen as a factor that could bind to a motif from the FGF-2 promoter (Ueba *et al*, 1999).

Since the earliest reports on MBD1 various isoforms of this protein have been identified both in human (Fujita *et al*, 1999) and in mouse (Hendrich and Bird, 1998; Jørgensen *et al*, 2004). These isoforms share a common N-terminal MBD as well as a C-terminal TRD (discussed in section 1.5.2), but differ in terms of which zinc finger domains are present and in other sequences at the C-terminus.

1.5.2 DNA binding and transcriptional repression by MBD1

Recombinant full-length human MBD1 (isoform PCM1) was shown to bind specifically to methylated DNA and to repress transcription in a methylation-dependent fashion *in vitro* (Cross *et al*, 1997). Recombinant mouse MBD1 was similarly shown to bind to DNA in a methylation-dependent manner *in vitro* (Hendrich and Bird, 1998).

Transiently expressed human MBD1 was shown to repress transcription from transfected reporter genes (Fujita *et al*, 1999). In the case of the isoforms MBD1v3 and MBD1v4 this was a methylation-dependent phenomenon. However, in the case of MBD1v1 and MBD1v2 repression of expression occurred irrespective of whether it was from a methylated promoter and the ability to repress unmethylated reporters depended on the presence of the zinc finger domain CxxC3 (Fujita *et al*, 2000). Consistent with this, the isolated CxxC3 domain from mouse Mbd1 binds specifically to unmethylated CpG containing DNA *in vitro* (Jørgensen *et al*, 2004). Furthermore, the localization of over-expressed MBD1 to mouse pericentric heterochromatin requires an intact MBD in wild-type cells, whereas an intact CxxC3 is needed for this to occur in DNA methylation-deficient cells. The ability of MBD1 to interact with either methylated or unmethylated CpGs via distinct domains should not be unexpected given that it was independently discovered in a yeast one hybrid screen (where DNA methylation is absent) and by sequence similarity to MeCP2. It remains to be determined whether *in vivo* MBD1 can interact simultaneously with modified and unmodified DNA or if these two modes of binding are mutually exclusive.

PCM1 was shown to repress transcription from transfected reporter genes when tethered via a heterologous DNA binding domain and this assay allowed the mapping of a transcriptional repression domain (TRD) to the C-terminus of MBD1 (Ng *et al*, 2000). This initial study reported MBD1-mediated repression to be sensitive to TSA, suggesting that histone deacetylases might be involved in this silencing. However, a different group contradicted this finding and instead proposed that an interaction between MBD1 and MCAF (MBD1-containing chromatin-associated factor) is involved in repression (Fujita *et al*, 2003a). This interaction was uncovered in a yeast two-hybrid screen using the TRD of MBD1 as bait and it was verified to occur both *in vitro* and *in vivo* by co-immunoprecipitation of endogenous proteins and co-localization of over-expressed proteins. Furthermore, mutations in the TRD that disrupt this interaction were found to interfere with repression and over-expression of MCAF stimulated MBD1-mediated repression in reporter assays.

There is also evidence that MBD1 can repress endogenous genes. In HeLa cells the methylated CpG island promoter of p53BP2 was identified as an MBD1 binding site by cloning DNA fragments isolated by chromatin immunoprecipitation (Sarraf and Stancheva, 2004). This promoter, which is ordinarily silenced in HeLa cells can be reactivated either by treatment with 5-azadeoxycytidine or by siRNA-mediated depletion of MBD1. This study also found a direct interaction between MBD1 and the histone H3 lysine K9 methyltransferase SETDB1, and showed that repression was SETDB1 dependent. Additionally the presence at the p53BP2 promoter of SETDB1 together with the trimethylation of H3-K9, a modification frequently associated with silent heterochromatin (Peters *et al*, 2002), was found to depend upon MBD1. It is possible that this report together with the work on MCAF revealed different aspects of the same repressive mechanism as MCAF has been reported to stably associate with SETDB1 and stimulate its catalytic activity (Wang *et al*, 2003). Moreover, transient expression of SETDB1 in addition to MCAF further enhanced MBD1 mediated repression of a reporter gene (Ichimura *et al*, 2005).

MBD1 has also been shown to interact with the p150 subunit of the chromatin assembly factor CAF-1 (Reese *et al*, 2003; Sarraf and Stancheva, 2004). Methylated DNA blocks this interaction *in vitro*, and *in vivo* it is only observed during S-phase where it depends upon the presence of active replication forks. Also reported was an S-phase specific MBD1-dependent interaction between CAF-1 and SETDB1. Taken together these data led to the proposal of a model where a complex of MBD1 and SETDB1 is displaced from methylated DNA during S-phase and is thus recruited to CAF-1 in order that nucleosomes are modified at on H3-K9 as they are deposited at a promoter silenced by DNA methylation and MBD1. This was born out by the observation that following transient depletion of MBD1 the recovery of H3-K9 trimethylation at the p53BP2 promoter was dependent on cell division (Sarraf and Stancheva, 2004). Consistent with a link between CAF-1 and MBD1 an earlier study characterized the interaction between these two proteins and showed that over-expression of the C-terminus of CAF-1 p150

led to the mislocalization of MBD1 away from the otherwise still intact pericentric heterochromatin (Reese *et al*, 2003).

The generation and analysis of *Mbd1*-null mice supported the view of Mbd1 as a transcriptional repressor only to a limited degree. These mice showed mild neurological phenotypes including lessened neurogenesis and decreased long term potentiation together with autism-like behavioural defects (Zhao *et al*, 2003; Allan *et al*, 2008). Consistent with this, mutations in human MBD1 have been reported in human patients with autism but the number of people examined in this study was low (Li *et al*, 2005). To date, however, the molecular basis for the defects seen in the absence of *Mbd1* remains unclear. Initial microarray based expression analysis of the hippocampi from *Mbd1*-null mice revealed few changes except for up-regulation of the intra-cisternal A particle (IAP) transcript (Zhao *et al*, 2003). A later study found a two-fold increase in the expression level of a serotonin receptor Htr2c in these animals (Allan *et al*, 2008). However, in both cases it is unclear if the effects are a direct consequence of the removal of MBD1, and if so, to what extent these transcriptional changes contribute to the phenotype. It remains possible that a more striking role for MBD1 as a transcriptional repressor will emerge in tissues other than the brain, perhaps under conditions not yet investigated.

1.5.3 Regulation of MBD1 by SUMO modification

The yeast two-hybrid screen that identified SETDB1 and the p150 subunit of CAF-1 as interaction partners of MBD1 also identified the SUMO (small ubiquitin-like modifier) E3 ligases PIAS1 and PIAS3 as putative MBD1 binding partners (Lyst *et al*, 2006). In mammals the SUMO family of proteins comprises SUMO-1, SUMO-2 and SUMO-3 (Hay, 2005). The C-terminal diglycine motifs of these small proteins can be coupled to lysine residues within target proteins via an isopeptide bond in a fashion reminiscent of ubiquitin modification (Hay, 2005). SUMO modification will be discussed in more depth in chapter five.

A pool of MBD1 in HeLa cells was shown to be modified with SUMO-1 on lysine 450 and lysine 489. Over-expression or depletion of the PIAS proteins identified as MBD1 binding partners led respectively to an increase or a decrease in the level of SUMO modification of MBD1 (Lyst *et al*, 2006). SUMO modification of MBD1 by PIAS protein over-expression was further shown to disrupt its interaction with SETDB1 *in vivo* leading to loss of trimethylation of H3-K9 from the p53BP2 promoter and expression of this MBD1 target gene (Lyst *et al*, 2006). The disruption of the interaction between MBD1 with SETDB1 is unlikely to be direct as SUMO modification of recombinant MBD1 does not disrupt its interaction with recombinant SETDB1 (Lyst *et al*, 2006). To date no circumstances have been identified when a physiological change in gene expression is mediated via SUMO modification of MBD1.

An independent study also documented SUMO modification of MBD1 (Uchimura *et al*, 2006). In addition to SUMO-1 modified MBD1 this report uncovered conjugation to SUMO-2 or SUMO-3 or both SUMO-2 and SUMO-3 (the antibodies employed could not distinguish these highly similar proteins). Furthermore, evidence was presented that SUMO modification promotes the association of MBD1 with MCAF *in vitro* and also that SUMO is required for the recruitment of heterochromatic factors such as HP1β to MBD1 containing foci *in vivo*. In contrast to the work on MBD1 and PIAS proteins these experiments suggest that SUMO modification might be expected to facilitate, rather than antagonize, transcriptional repression. The exact reasons for these apparent discrepancies are at present unclear, although they may be in part due to the different cell lines used in these studies.

1.5.4 Aims of this thesis

The MBD1 binding proteins described in previous sections are only some of the ones reported in the literature. Some other interactions are supported by quite substantial evidence, but the physiological meaning of these associations is always unclear. Conversely, others may suggest how MBD1 could function as a transcriptional

repressor, but often the evidence for these interactions is flimsy. An example of an interaction in the first category is that of MBD1 with the DNA repair protein methylpurine DNA glycosylase (MPG) (Watanabe *et al*, 2003). The association of these proteins is supported by multiple lines of evidence including yeast two-hybrid, co-localization of fluorescent protein fusions and co-immunoprecipitation of endogenous proteins (Watanabe *et al*, 2003). However, the functionality of this interaction remains completely unknown. Interactions in the second category include MBD1 binding to polycomb components (Sakamoto *et al*, 2007), HDAC3 (Villa *et al*, 2006) and heterochromatic proteins such as HP1 and Suv39h1 (Fujita *et al*, 2003b). These reports all suggest plausible mechanisms by which MBD1 could repress transcription. However, the evidence presented for the interactions concerned is always limited to some combination of *in vitro* experiments and analysis of over-expressed proteins *in vivo*.

Despite numerous reported interaction partners for MBD1, and in contrast with the situations for MBD2 and MeCP2, which are stably associated with the NuRD complex and monomeric in nuclear extracts, respectively (Zhang *et al*, 1999, Feng and Zhang 2001, Le Guezennec *et al*, 2006; Klose and Bird, 2004), it remained untested whether or not MBD1 formed part of a stable multi-subunit complex. If such a complex existed then its purification would allow the identification of its components. These would represent novel and likely informative MBD1 binding partners, or else they would help to determine which of the many reported MBD1 binding partners are most physiologically relevant by determining which ones associate with MBD1 in the greatest abundance and in the most stable fashion. Therefore a starting point of this work was to determine whether or not MBD1 is part of a stable multi-subunit complex, and if so, to purify it and identify its components.

2. Chapter Two - Materials and Methods

2.1 Bacterial methods

2.1.1 Bacterial culture conditions

Escherichia coli (DH5α and BL21(DE3) strains) were grown routinely at 37˚C in Luria-Bertani (LB) broth or on LB agar. Cells with plasmids were selected using antibiotics in the media (100 μg/ml ampicillin, 50 μg/ml kanamycin, 34 μg/ml chloramphenicol).

2.1.2 Preparation of competent cells and transformation

E.coli were streaked out onto LB agar and incubated overnight at 37˚C. A single colony was grown overnight at 37˚C in 2 ml SOB media (2% (w/v) tryptone (Bacto), 0.5% (w/v) yeast extract (Bacto), 10 mM NaCl, 2.5 mM KCl, 5 mM $MgSO_4$, 5 mM $MgCl_2$). The culture was transferred to 200 ml SOB media and grown at 20˚C until the optical density at 600 nm reached 0.5 before being cooled on ice for 10 min and centrifuged at 3000 g for 15 min at 4˚C. The cells were resuspended in 67 ml TB (10 mM PIPES-KOH pH 6.75, 15 mM $CaCl_2$, 0.25 M KCl, 55 mM $MnCl_2$), incubated on ice for 10 min and then centrifuged again. Finally the cells were resuspended in 16 ml TB, DMSO was added drop-wise to a concentration of 7% (v/v) and after a further 10 min on ice the cells were made into aliquots before being snap frozen in liquid nitrogen and stored long-term at -80˚C. For transformation either 1 ng plasmid DNA or 10 μl ligation reaction product were mixed with 75 μl of thawed competent cell suspension and incubated on ice for 10 min. The cells were then heat shocked at 42˚C for 2 min and then returned to ice for 10 min. 0.5 ml of LB media was added to the cells, which were then incubated with shaking at 37˚C for 1 h. Cells were then transferred to LB agar or LB broth containing selective antibiotics and incubated overnight at 37˚C.

2.1.3 Plasmid preparation

Plasmid DNA was purified from *E.coli* using a Quiagen miniprep kit except where the DNA was to be used for transfection of mammalian cells, in which case either a Quiagen midiprep or maxiprep kit was employed.

2.1.4 Protein expression and cell lysis

Plasmids were introduced into BL21(DE3) *E.coli* as described above. A 5 ml overnight culture was diluted into a larger volume (typically 1 litre) of antibiotic-containing LB broth and grown at 37˚C until the optical density at 600 nm reached 0.5. Expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the cells were then grown for 4 h at 30˚C. Cells were then collected by centrifugation for 10 min at 5000 g at 4˚C, washed in PBS, collected again, and then cell pellets were frozen and stored at -80˚C. Pellets from 1 litre cultures were resuspended in 25 ml ice cold bacterial lysis buffer (NE1 supplemented with 300 mM NaCl and 1 mM DTT replaced by 13 mM β-mercaptoethanol) and disrupted by 3 min sonication on ice (output 4 at duty cycle 30% on a Branson Sonifier 250). Lysates were cleared by 20 min centrifugation at 12,500 g at 4˚C followed by passing through a 0.2 μm pore syringe filter device.

2.2 Molecular cloning

2.2.1 Preparation of cDNA

Cell pellets were resuspended in 1 ml per 2 million cells of Trizol® reagent (Invitrogen) and left to stand for 2 min at room temperature before the addition of 1/5 volume of chloroform. The mixture was vortexed and allowed to stand for a further 2 min. After 20 min centrifugation at 14,000 g at 4˚C the top layer was taken and an equal volume of isopropanol was added. The RNA was precipitated for 1 h at -20˚C before being spun

down at 14,000 g for 30 min at 4˚C. The pellet was washed with ice cold 70% ethanol before being dried and dissolved in RNAse-free water (containing 2 mM DTT and 0.8 u/µl Ribolock (Invitrogen)). 4 µg RNA and 1.5 µg oligo(dT) were then combined in 16 µl RNAse free water (with 2 mM DTT and 0.8 u/µl Ribolock). After 10 min incubation at 65˚C the sample was set on ice and then added to a mixture of 2 µl 10 mM dNTPs (Fermentas), 6 µl first strand buffer (Invitrogen), 4 µl 75 mM DTT and 0.8 u/µl Ribolock. The mixture was incubated at 42˚C for 1 min then 2 µl of reverse transcriptase (Superscript II, Invitrogen) was added and the sample was left overnight at 42˚C.

2.2.2 Polymerase chain reaction

Polymerase chain reaction (PCR) was carried out in NEB buffer 4 supplemented with 3 mM $MgCl_2$, 10 mM dNTPs, 0.25 µM forward primer and 0.25 µM reverse primer. For a 100 µl reaction 5 units of Taq polymerase (prepared by Dr Jose de las Heras) and 0.5 units of Pfu DNA polymerase (Fermentas) were used. Template DNA was 1 ng plasmid DNA or serial dilutions of cDNA. The annealing temperature, elongation time, and number of cycles of PCR were varied according to the template DNA, product length and primers. In some instances, coding sequences for fusion proteins were constructed by PCR-mediated recombination. Initial PCR reactions were carried out where the reverse primer of one reaction and the forward primer of the other reaction carry overhangs such that the primers are complementary to each other. A second PCR used the products of the initial reactions as templates such that the sequences are joined in the final product. Restriction sites were included in primer sequences where appropriate.

2.2.3 Restriction digests

Restriction digests were carried out using enzymes from fermentas or NEB (New England Biolabs) either singly or in combination using the buffer, incubation time and

temperature recommended by the manufacturer. For cloning purposes 5 µg of plasmid DNA would be digested as well as 1 µg purified PCR product.

2.2.4 Purification of restriction fragments and PCR products

PCR products were purified using a Quiagen PCR purification kit. Restriction fragments were subjected to agarose gel electrophoresis (1% agarose in TAE), visualised under UV by ethidium bromide staining, excised and purified with a gel extraction kit (Eppendorf).

2.2.5 Determination of DNA concentration and quality

Nucleic acids were quantified based on their absorbance at 260 nm using a NanoDrop 1000 (Thermo Scientific). The same instrument was used to verify that DNA used for transfections had an A260/280 ratio greater than 1.8.

2.2.6 Dephosphorylation and ligation of DNA

Whenever compatible ends were generated by restriction digestion of vector DNA the sample would subsequently have its 5' phosphates removed prior to insert ligation. To this end digested and purified vector DNA would be incubated for 30 min at 37˚C with 5 units of Antarctic Phosphatase (NEB) in the buffer provided by the manufacturer. The enzyme would then be heat inactivated at 65˚C for 10 min. Ligations were carried out overnight at 16˚C using a 3:1 molar ratio of insert to vector DNA and 5 units of T4 DNA ligase (Fermentas) in 10 µl reactions using the buffer supplied by the manufacturer.

2.2.7 DNA sequencing

DNA sequencing was carried out using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Roche) according to the manufacturer's instructions. Reactions were then analysed

on an ABI sequencer by the ICAPB sequencing service at the University of Edinburgh. A complete list of plasmids generated in this study can be found in appendix I.

2.3 Mammalian cell culture

2.3.1 Culture conditions

HeLa cells, Phoenix cells and mouse embryonic fibroblasts were grown routinely at 37˚C under 5% CO_2 in Dulbecco's Modified Eagle's Medium (Sigma) supplemented with 10% foetal bovine serum (Perbio) and penicillin-streptomycin-glutamine (Invitrogen). Cells were passaged by washing with Dulbecco's PBS (Invitrogen) followed by treatment with 0.05% Trypsin and 0.53 mM EDTA (Invitrogen) in D-PBS to detach them from the flask. The reaction was stopped by adding one volume of serum containing growth medium. Cells were then diluted into fresh medium in a new flask or harvested by centrifugation for 5 min at 250 g followed by washing with D-PBS. Cell pellets were stored at -80˚C for future analysis.

2.3.2 Cryogenic storage

Cells were harvested using trypsin, centrifuged and then resuspended in 1 ml of growth medium supplemented with 20% DMSO per 5 million cells. Cells were frozen at -80˚C before transfer to liquid nitrogen for long term storage. Stored cells were defrosted by warming to 37˚C, washed in growth medium and seeded into a new culture vessel.

2.3.3 Transfection and transduction

Cells were transfected using the JetPei reagent (Autogen Bioclear) according to the manufacturer's instructions. Retroviruses for transduction were produced by transfection

of Phoenix cells with pBABE vectors. The media was changed 24 h post transfection and the supernatant after 48 h was taken as viral stock. Serial dilutions of this were then applied to HeLa cells for 24 h to achieve infection. During infection the growth medium was supplemented with 5 μg/ml polybrene (Sigma).

2.3.4 Preparation of nuclear extracts

All steps were carried out at 4˚C unless otherwise stated. Cell pellets were resuspended in 2ml of nuclear extraction buffer NE1 (20 mM HEPES-KOH pH 7.5, 10 mM KCl, 1 mM $MgCl_2$, 0.1% Triton-X-100, 1 mM DTT and protease inhibitors (Sigma P8340)) per 50 million cells and allowed to swell on ice for 5 min. The cells were disrupted with 10 strokes of a Dounce homogenizer and nuclei were collected by centrifugation at 500 g for 5 min. In some cases the nuclei were then treated with 10 units (per million nuclei) benzonase (Merck) for 5 min at room temperature. This involved resuspending the nuclei in as small a volume of NE1 as possible before adding the enzyme. The nuclei were then resuspended in an appropriate volume of nuclear extraction buffer NE2 (NE1 supplemented with 420 mM NaCl) such that the final concentration of NaCl was 300 mM. In some cases the nuclei were resuspended in NE1 buffer supplemented with less than 420 mM NaCl such that the final salt concentration would be 150 mM. Finally the suspension was incubated with mixing for 40 min before being centrifuged for 20 min at 14,000 g. The supernatant was taken as nuclear extract.

2.3.5 Cell fixation and microscopy

Cells expressing GFP fusion proteins were grown on coverslips, washed in D-PBS (Invitrogen) and then fixed by incubating in 4% paraformaldehyde in PBS for 10 min at room temperature. The cells were then permeabilized by incubation in 0.25% Triton-X-100 in PBS for 3 min at room temperature. The coverslips were then mounted on slides using Vectashield® mounting medium containing 4',6-Diamidine-2'-phenylindole dihydrochloride (DAPI) and stored in the dark before analysis. Finally the cells were

observed using an Olympus BX61 microscope and a ColorViewII camera. The analySIS software package was used for image capture.

2.4 Protein purification

2.4.1 Purification of proteins containing a 3xFLAG-tag

FLAG M2 agarose (Sigma) was equilibrated with the buffer used for nuclear extraction. Extracts were then mixed with the resin for 1 hour at 4˚C before being washed four times with nuclear extraction buffer. Bound proteins were then eluted by mixing for 4 h at 4˚C with this buffer containing 750 µg/ml 3X FLAG peptide (Sigma). Alternatively proteins were eluted by boiling the resin in Laemmli buffer.

2.4.2 Immobilized metal affinity chromatography (IMAC)

Chelating sepharose (Amersham) was charged with nickel ions according to the manufacturer's instructions and then equilibrated with bacterial lysis buffer supplemented with 10 mM imidazole. Bacterial lysates or nuclear extracts were adjusted to 10 mM imidazole and incubated with the resin for 1 h at 4˚C with mixing. The resin was then poured into a disposable chromatography column (Biorad) and washed four times with twenty column volumes of wash buffer (bacterial lysis buffer supplemented 20 mM imidazole). Proteins were then eluted by washing the resin five times with one column volume of the same buffer supplemented with 250 mM imidazole.

2.4.3 Purification of proteins containing a GST-tag

Mixtures containing a protein with a GST-tag were incubated for 1 h with mixing at 4˚C with glutathione sepharose (Amersham) pre-equilibrated with nuclear extraction buffer supplemented with 300 mM NaCl. After binding the resin was poured into a disposable

chromatography column (Biorad) and washed four times with twenty column volumes of nuclear extraction buffer. Proteins were then eluted by washing the column ten times with one column volume of elution buffer (50 mM Tris-HCl pH 8, 20 mM glutathione, 300 mM NaCl, 0.1% Triton-X-100, 1 mM DTT). Eluted proteins were dialysed back into nuclear extraction buffer, made into aliquots and then stored at -80˚C.

2.4.4 Cation exchange chromatography

For an extract from a 1 litre bacterial culture expressing MBD1, 1 ml of SP sepharose (Amersham) was equilibrated with bacterial lysis buffer. The resin was then incubated with the clarified bacterial lysate with tumbling for 1 h at 4˚C before being applied to a disposable chromatography column. The resin was washed three times with ten column volumes of bacterial lysis buffer supplemented with 500 mM NaCl. Proteins were eluted with five 1 ml washes with this buffer supplemented with 1 M NaCl.

2.5 Protein analysis

2.5.1 Determination of protein concentration

Concentrations of partially purified proteins were estimated by SDS-PAGE followed by Coomassie staining using serial dilutions of the test protein and known amounts of a BSA (bovine serum albumin) standard. Protein concentrations in extracts were estimated by measuring absorbance at 280 nm using a NanoDrop 1000 (Thermo Scientific) and with the BCA (bicinchoninic acid) reagent (Pierce).

2.5.2 SDS-PAGE

Gels were cast and subjected to electrophoresis in the Mini-PROTEAN apparatus (Bio-Rad). The upper stacking gel was composed of 125 mM Tris-HCl pH 6.8, 4% (w/v)

acrylamide/bis-acrylamide (37.5:1), 0.1% SDS, 0.05% ammonium persulfate (APS), 0.2% TEMED. The lower separating gel was made of 375 mM Tris HCl pH 8.8, 7-20% (w/v) acrylamide/bis-acrylamide (37.5:1), 0.1% SDS, 0.05% APS, 0.2% TEMED. The running buffer was 25 mM Tris, 192 mM glycine, 0.1% SDS. Gels were run at 270 V. Prior to loading samples were boiled in Laemmli buffer (50 mM Tris-HCl pH 6.8, 100 mM DTT, 2% SDS, 10% glycerol, 0.1% bromophenol blue).

2.5.3 Coomassie staining

After electrophoresis gels were rinsed briefly in water before being incubated in Coomassie staining solution (0.2% Coomassie (Sigma) (w/v), 50% methanol (v/v), 10% acetic acid (v/v)) for 1 h. Gels were then destained to reveal protein bands by repeated incubations and agitation in 50% (v/v) methanol with 10% (v/v) acetic acid. Finally gels were washed in water, wrapped in cellophane and dried.

2.5.4 Western blotting

After separation by SDS-PAGE proteins were transferred to 0.4 μm pore nitrocellulose membranes (Bio-Rad) using the Mini Trans-Blot® module (Bio-Rad). The transfer buffer used was 25 mM Tris, 192 mM glycine. Transfers were carried out for 1 h at 400 mA. Membranes were stained with Ponceau S solution (2% Ponceau S (w/v), 30% trichloroacetic acid, 30% sulfosalicyclic acid) to ensure equal loading of lanes before being washed with PBS supplemented with 0.1% Tween-20. Membranes were incubated in blocking solution (4% fat free milk powder in TBS) for 30 min at room temperature. Primary antibodies were applied to membranes in blocking solution overnight at 4˚C typically at a concentration of 0.5 μg/ml. Primary antibodies used in Western blots were as follows: anti-MBD1 (IMG-306 from Imgenex for human and M254 from Santa Cruz for mouse), anti-SETDB1 (Upstate), anti-HDAC1 (sc-7872), HDAC3 (Active Motif), anti-G9a (Sigma), anti-GLP (MBL international), anti-EZH2 (a gift from Klaus Hansen),

anti-ACF1 (a gift from Patrick Varga-Weisz), anti-LSD1 (Bethyl), anti-GFP (CRUK), M2 anti-FLAG (Sigma) and HRP-conjugated anti-GST (Amersham). The membrane would then be washed with PBS supplemented with 0.1% Tween-20 before application of the HRP-conjugated secondary antibody (Sigma). After washing again with PBS supplemented with 0.1% Tween-20, bands were visualized by applying developing solution (100 mM pH 8.8, 0.01% H_2O_2, 1.5 mM luminol, 0.25 mM p-coumaric acid) to the membrane and then exposing it to X-ray film.

2.5.5 Immunoprecipitation

All steps were carried out at 4˚C unless otherwise stated. Nuclear extracts (typically 100 µl at 10 mg/ml protein concentration) containing GFP fusion proteins were added to 2 µg of monoclonal antibody against GFP (CRUK) and incubated with mixing for 4 h. The mixture was added to 5 µl of protein G sepharose (Amersham) pre-equilibrated with nuclear extraction buffer and capture of immunocomplexes was allowed to proceed for 1 h with mixing. The protein G sepharose was then washed four times with 0.5 ml nuclear extraction buffer then bound proteins were used in enzymatic assays or eluted by boiling in Laemmli buffer for analysis by Western blotting.

2.5.6 SUMO protease assay

Substrates were mixed with sources of protease activity and incubated for 1 h at 37˚C in nuclear extraction buffer supplemented with 300 mM NaCl. The catalytic domain of the yeast SUMO protease Ulp1p was obtained from Invitrogen and used in accordance with the manufacturer's instructions except the buffer conditions were the same as for the other SUMO protease assays. When nuclear extracts or fractions were the source of protease activity between 1 µg and 30 µg total protein were employed. Typically around 50 ng of substrate were used. The reaction was then assayed by SDS-PAGE followed by Western blotting with antibodies against GST or against MBD1.

2.5.7 Gel filtration

A 10 mm by 30 cm Superose 6 HR 10/30 column (Amersham) was calibrated with proteins of known Stoke's radii (Amersham gel filtration calibration kit). The column was equilibrated with nuclear extraction buffer supplemented with 200 mM NaCl before loading the sample (typically just less than 0.5 ml) onto the column. The flow-rate was 0.3 ml/min and 0.64 ml fractions were collected. The whole procedure was carried out at 4˚C. Fractions were then analysed by SDS-PAGE followed by Western blotting.

2.5.8 Sucrose gradient sedimentation

Sucrose gradients were made in nuclear extraction buffer supplemented with 200 mM NaCl. Linear 2 ml gradients contained 5-20% (w/w) sucrose and were made by layering 0.5 ml of sucrose solutions of decreasing concentrations on top of each other before being left to stand for 4 h at 4˚C. A sample of 150 µl would be layered on top of the gradient before centrifugation at 4˚C for 4 h at 368,000 g in an SW 55 Ti swinging bucket rotor. Fractions of 165 µl were collected from the top of the gradient and analysed by SDS-PAGE followed by Western blotting or Coomassie staining. For calibration purposes, identical gradients were run in parallel containing proteins of known sedimentation coefficients (Amersham gel filtration calibration kit).

2.5.9 Mass spectrometry

Proteins to be analysed were run approximately 1 cm by SDS-PAGE, stained with Coomassie and excised. Mass spectrometry was carried out by Flavia Alves in the laboratory of Juri Rappsilber as described elsewhere (Milligan *et al*, 2008) except that database used was IPI-Human. Analysis of post-translational modifications on histones H3 and H4 was carried out in the laboratory of Axel Imhof as described elsewhere (Loyola *et al*, 2006).

3. Chapter Three – Evidence for self association of MBD1

3.1 Endogenous MBD1 does not behave as a monomer

It was previously shown that MBD1 in HeLa nuclear extracts has a Stoke's radius of approximately 5.4 nm as assessed by size exclusion chromatography (Ng *et al*, 2000). However, in Western blots the MBD1 antibody used in this study recognised a doublet in HeLa nuclear extracts with the faster migrating of the two bands corresponding closely to the migration of recombinant hexahistidine-tagged MBD1 expressed in *E.coli* (isoform PCM1). In contrast the monoclonal MBD1 antibody used in subsequent studies and validated by siRNA depletion experiments (Sarraf and Stancheva, 2004) recognises bands of lower molecular weight than recombinant PCM1 expressed in either *E.coli* or exogenously in HeLa cells (see section 5.7). This result cast doubt on the specificity of the antibody used in the earlier size exclusion analysis of MBD1. It was therefore decided to re-visit this observation as well as carrying out a more thorough biophysical analysis of the endogenous MBD1 in HeLa cells.

In gel filtration the endogenous MBD1 protein in HeLa nuclear extracts eluted from a Superose 6 column in the same fashion as the apoferritin 443 kDa size standard (Figure 3.1A) indicating a Stoke's radius of approximately 6.1 nm. By analysing the same extracts on 5-20% sucrose gradients a sedimentation coefficient of S = 5.4 was determined (Figure 3.1B). Given the Stoke's radius and sedimentation coefficient it is possible to calculate the molecular mass of a protein or protein complex even when present in a crude extract (Siegel and Monty, 1966). This is achieved by substituting these values into the following equation derived by Siegel and Monty:

$$M_r = 6\pi\, \eta_{20,w} \bullet s_{20,w} \bullet R_S \bullet N / (1 - \rho_{20,w}\, v)$$

In this equation R_S = Stoke's radius (cm), $s_{20,w}$ = sedimentation velocity (S x 10^{-13}), $\eta_{20,w}$ = viscosity of water at 20 ˚C (0.01002 g/cm/s), N = Avogadro's number (6.022 x 10^{23} mol^{-1}), $\rho_{20,w}$ = density of water at 20 ˚C (0.9981 g/cm^3), and v = partial specific volume used (0.725 cm^3/g). The application of this equation to the native MBD1 protein found in HeLa cells led to the calculation of a molecular mass of approximately 136 kDa. As the theoretical molecular mass of this protein is only 61 kDa it can be concluded that MBD1 either self associates or is stably bound to another factor in HeLa nuclear extracts.

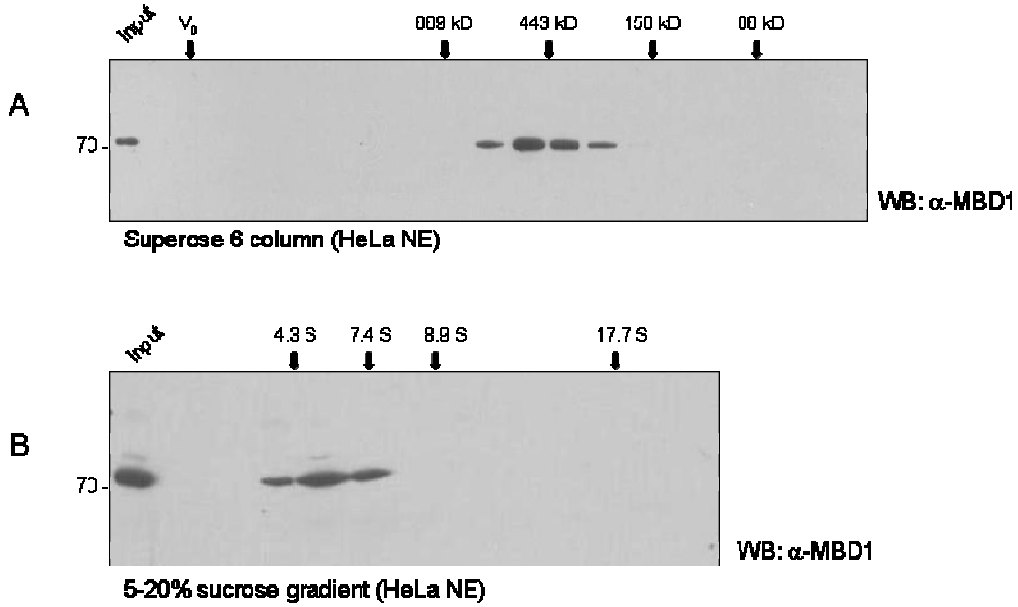

Figure 3.1 – Native MBD1 in HeLa nuclear extracts is not monomeric

(A) Western blot using antibodies against MBD1 to analyse the elution of the native protein in HeLa nuclear extracts from a Superose 6 gel filtration column. A Stoke's radius of 6.1 nm was measured.
(B) Western blot analysis of migration of native MBD1 from HeLa extracts in 5-20% sucrose gradients. A sedimentation coefficient of S = 5.4 was measured.

3.2 Bacterially expressed MBD1 behaves as a monomer

A similar biophysical analysis was carried out using untagged bacterially expressed MBD1, which had been partially purified using cation exchange chromatography. Gel filtration analysis revealed a Stoke's radius of 5.4 nm (Figure 3.2A) and sucrose gradients revealed a sedimentation coefficient of 2.6 S (Figures 3.2B). From these values a mass of 58 kDa was calculated using the Siegel and Monty equation given above. This is in close agreement with the theoretical molecular mass of MBD1 of 61 kDa and so, it appears that recombinant MBD1 behaves as a monomer suggesting that the protein does not strongly interact with itself.

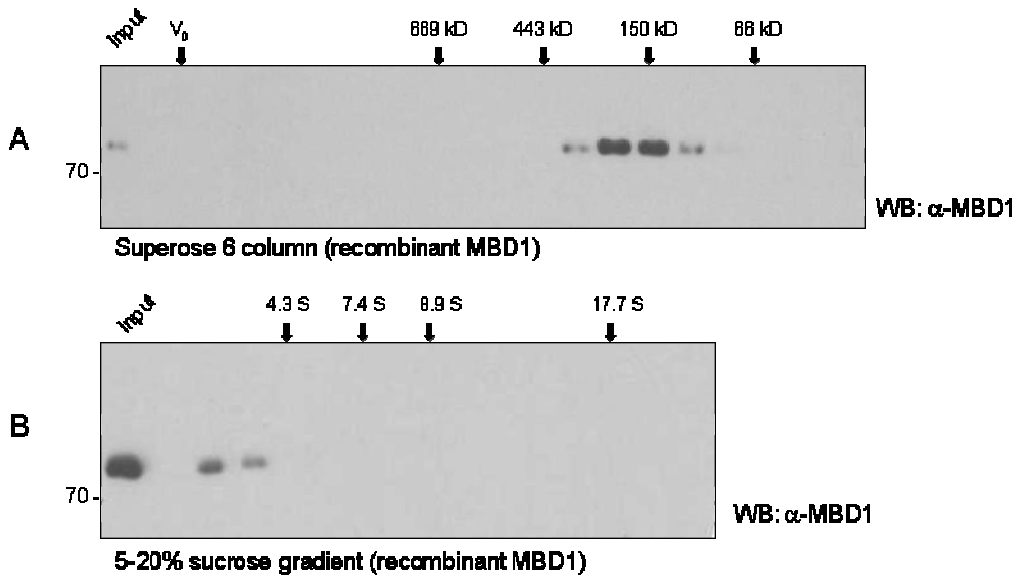

Figure 3.2 – Recombinant MBD1 expressed in *E.coli* behaves as a monomer

(A) Size exclusion analysis using a Superose 6 gel filtration column of recombinant MBD1 partially purified from *E.coli*. The fractions were analysed by Western blotting and a Stoke's radius of 5.4 nm was measured.
(B) Bacterially expressed MBD1 was fractionated on 5-20% sucrose gradients. Western analysis of the fractions revealed a sedimentation coefficient of S = 2.6 for this protein.

A plausible explanation for the differential behaviour of native MBD1 compared to the recombinant protein could be the stable interaction of the former with a binding partner. However, this remained only a tentative hypothesis as incorrect folding of the recombinant protein could not be ruled out, and so self association of the native protein was still a possibility.

3.3 SETDB1 association does not explain the apparent size of MBD1

MBD1 has previously been reported to form a stable complex with the histone H3 lysine K9 methyltransferase SETDB1 with the majority of MBD1 in HeLa nuclear extracts being involved in this interaction (Sarraf and Stancheva, 2004). MBD1 has also been reported to interact with MCAF (Fujita *et al*, 2003a) which is a co-factor of SETDB1 (Wang *et al*, 2003). Therefore association with one or both of these proteins might be expected to lead to an increased apparent size of MBD1 in HeLa nuclear extracts as assessed by gel filtration chromatography and sucrose gradient sedimentation. However, given the molecular masses of SETDB1 and MCAF being 143 kDa and 135 kDa respectively, it seems that these proteins are too large, even individually, to account for the size of MBD1 measured here. Furthermore, SETDB1 and MCAF are known to form hetero-multimers together making this problem even more pronounced (Wang *et al*, 2003).

Gel filtration analysis of SETDB1 in HeLa extracts was carried out here (Figure 3.3) and this confirmed the large apparent size of the protein that was previously noticed and explained by association with MCAF (Wang *et al*, 2003). Furthermore, it is evident that there is very limited, if any, overlap between the elution profiles of SETDB1 and MBD1 (Figure 3.3 compared with Figures 3.1A or 3.4). This makes the existence of a stable complex containing both of these proteins seem unlikely.

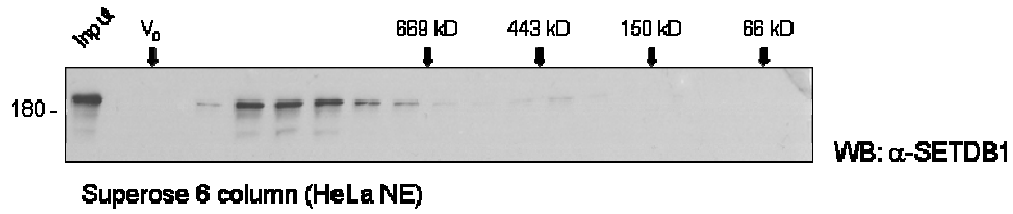

Figure 3.3 – Separation of MBD1 and SETDB1 by gel filtration

HeLa nuclear extracts were fractionated on a Superose 6 gel filtration column. Western blot using SETDB1 antibodies reveals the elution profile of this protein. SETDB1 has a greater Stoke's radius than MBD1 with little overlap between the two proteins.

3.4 Nucleic acid does not contribute to the apparent size of MBD1

Transcription of rDNA is regulated in part by the nucleolar chromatin remodelling complex NoRC. TIP5 (TTF-interacting protein 5), the largest subunit of NoRC, contains a TAM (TIP5/ARBP/MBD) domain which has similarity to the methyl binding domain. TIP5 is known to strongly associate with 150-300 nucleotide RNAs, which show complementarity to regulated sequences in the rDNA promoter (Mayer *et al*, 2006). This leads to the hypothesis that the MBD of MBD1 might be stably associated with an RNA molecule leading to the large apparent molecular mass of MBD1 in HeLa nuclear extracts. In order to test this possibility, isolated HeLa nuclei were treated with the promiscuous benzonase nuclease which contains both DNAse and RNAse activities. This reaction was carried out as described in chapter two except the incubation time was 1 h and the temperature was 37 ˚C. Proteins were then extracted from the nuclei in the usual way and the nuclear extract was fractionated on a Superose 6 gel filtration column. The Stoke's radius of native MBD1 in HeLa nuclear extracts was unchanged by this treatment indicating that stable association with RNA does not account for the large apparent size of MBD1 in this assay (Figure 3.4).

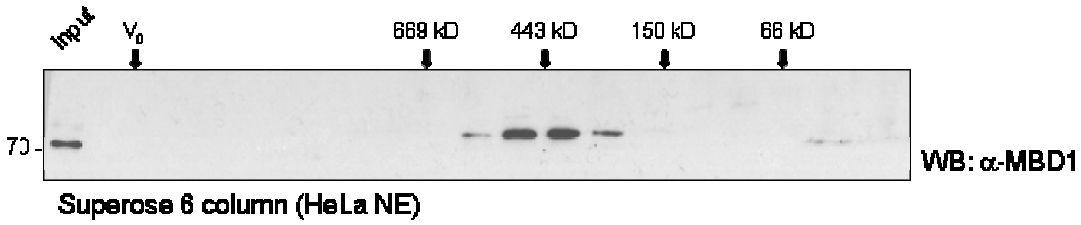

Figure 3.4 – Nucleic acid does not contribute to the large apparent size of MBD1

HeLa nuclear extracts treated with benzonase nuclease were fractionated on a Superose 6 gel filtration column and the fractions analysed by Western blot. MBD1 is found to elute in the same fractions as when the extract is not treated with benzonase.

3.5 Affinity tagging of MBD1

Given that at this stage neither self-association, binding of SETDB1 nor interaction with nucleic acid seemed likely to explain the apparent molecular mass of MBD1, the notion of a stable interaction with an as yet unidentified binding partner was entertained. Historically the biochemical fractionation of proteins has frequently led to the identification of stably associated polypeptides. For example, in the case of histone deacetylases, other components of the Sin3a and NuRD co-repressor complexes were found in this way (Zhang *et al*, 1997; Zhang *et al*, 1999). However, attempts to purify native MBD1 and identify associated proteins using conventional chromatography techniques were unsuccessful (data not shown). This was partly due to difficulties in eluting a large proportion of native MBD1 from many of the chromatography resins tested, making it impossible to achieve great enough enrichments. Nevertheless, this line of work found that even after a five column purification procedure, native MBD1 maintained its large apparent size as assessed by gel filtration. Therefore any binding partners of MBD1, which contributes to its size, are stably associated and the putative complex should be biochemically purifiable.

A more modern approach towards the isolation of protein complexes involves expression of a protein of interest as a fusion with tags that facilitate its purification. For example, purification of a tagged form of the histone H3 lysine K4 demethylase JARID1d revealed an interaction with Ring6a/MBLR which is a protein with homology to components of the polycomb complex PRC1 (Lee *et al*, 2007). Therefore, in order to investigate the nature of the putative MBD1-containing complex, a retroviral system was used to express this protein in HeLa cells as a fusion with a C-terminal S-tag (Terpe, 2003), streptavidin binding peptide (Wilson *et al*, 2001), histidine hexamer and GFP-tag (MBD1-SSH-GFP).

Fluorescence microscopy of HeLa cells expressing MBD1 tagged in this way revealed a diffuse localization with approximately six more strongly fluorescent foci per cell (Figure 3.5A). In contrast GFP fused to the affinity tags alone was localized more evenly throughout the cell (Figure 3.5B). The localization observed here for exogenous MBD1 is very similar to that observed for the endogenous protein by immunofluorescence (Sakamoto *et al*, 2007) (Figure 3.5C). It seemed therefore that the tagging and over-expressing of MBD1 in this fashion did not perturb its normal behaviour in this assay, and so purifications of this protein were undertaken.

Purification of MBD1-SSH-GFP, however, proved to be a fruitless endeavour. In spite of being readily detectable in the microscope, this protein was surprisingly absent from nuclear extracts as assessed by Western blotting with antibodies against MBD1. These apparently contradictory observations were resolved by the discovery that MBD1-SSH-GFP is not extracted from the nucleus under the same conditions as the endogenous protein, with MBD1-SSH-GFP extraction requiring the solubilisation of a nuclear pellet fraction by sonication (Figure 3.5D). The differential solubility of endogenous MBD1 and over-expressed MBD1-SSH-GFP is perhaps surprising in the light of the correct localization of the exogenous protein. However, given this crude biochemical distinction between the two entities, work with this construct was abandoned as it seemed unlikely that this fusion protein was a good model to use for the study of endogenous MBD1.

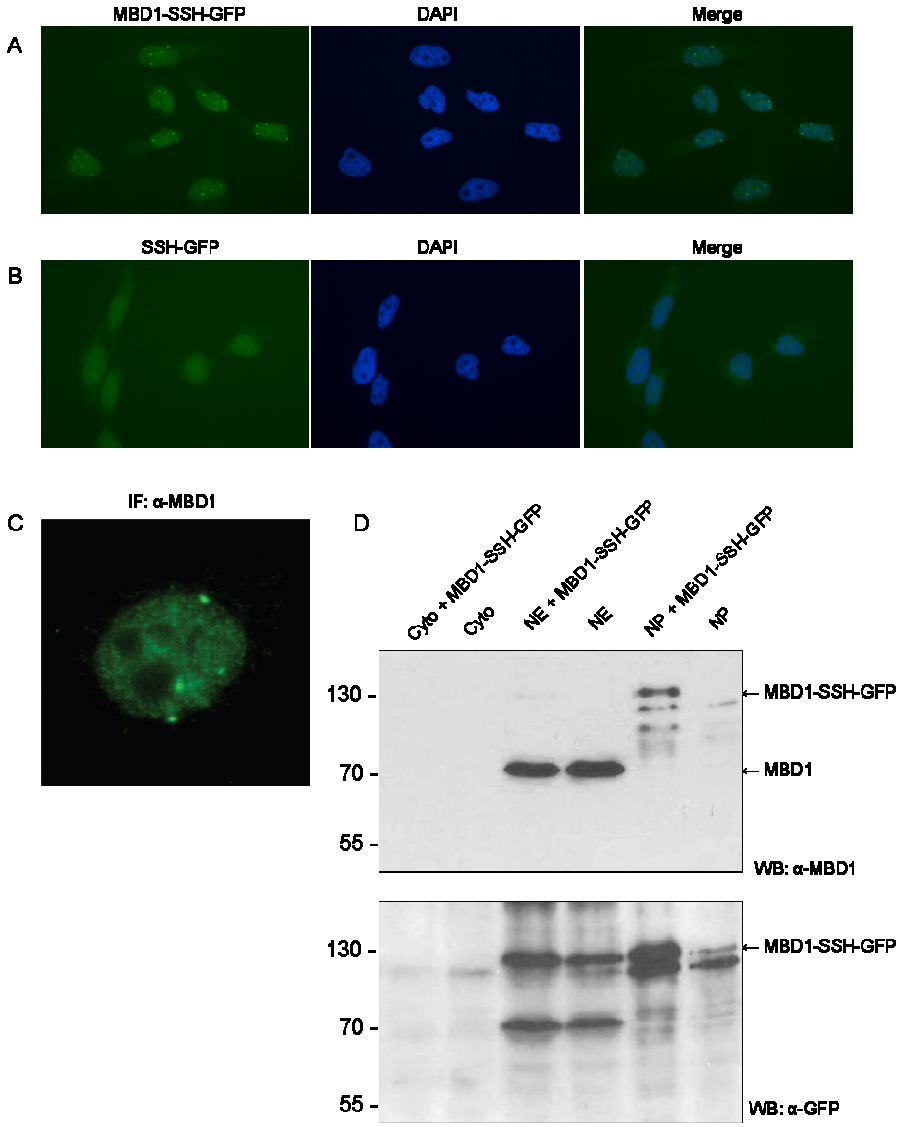

Figure 3.5 – Generation of affinity tagged MBD1

(A) Fluorescence microscopy of HeLa cells expressing MBD1-SSH-GFP and DAPI stained. MBD1 forms approximately six bright foci against a diffuse nuclear signal.
(B) Fluorescence microscopy image of HeLa cells expressing SSH-GFP and stained with DAPI. The protein distributes throughout the cell.
(C) Immunoflourescence (taken from Sakamoto *et al*, 2007) showing the localization of endogenous MBD1 in HeLa cells. The strong foci against the background of diffuse nuclear signal is similar to the pattern observed for MBD1-SSH-GFP.
(D) Western analysis of subcellular fractions of HeLa cells expressing MBD1-SSH-GFP using antibodies against MBD1 (upper panel) and GFP (lower panel). MBD1-SSH-GFP is found in the nuclear pellet fraction whereas the native protein is in the nuclear extract.

In order to try to overcome this problem MBD1 was expressed in HeLa cells with various other combinations of tags. Initially GFP was removed from the version of MBD1 described above whilst a 3xFLAG tag was added to its N-terminus, and this fusion protein was expressed under the control of a CMV (Cytomegalovirus) promoter by transient transfection. In contrast to the previous case, this protein could be extracted from the nucleus using 0.3 M sodium chloride. However, when these extracts were fractionated on a Superose 6 gel filtration column, this form of MBD1 was found to elute at the void volume indicating the formation of an aggregate (data not shown). Once again, therefore, a construct had to be discarded due to major biochemical differences between endogenous and exogenous MBD1. Finally MBD1 was similarly expressed under the control of a CMV promoter except this time with only N-terminal 3xFLAG and S-tags (FS-MBD1). This protein was readily extracted from the nucleus and when these extracts were fractionated by either sucrose gradient sedimentation (Figure 3.6A) or size exclusion chromatography (Figure 3.6B and 3.6C) both the Stoke's radius and S value of FS-MBD1 were found to closely match what was observed for the endogenous protein. Therefore the use of these tags appears not to alter these critical biochemical properties of MBD1 and so a potentially useful tool with which to study this protein had been generated.

3.6 Absence of stable and stoichiometric MBD1 binding partners

In order to try to identify any stable binding partners of FS-MBD1 the over-expressed fusion protein was immunoprecipitated from HeLa nuclear extracts using the M2 monoclonal antibody against the FLAG epitope. Following elution from the matrix using the 3xFLAG peptide purified proteins were separated by SDS-PAGE and then stained with Coomassie. Although purified FS-MBD1 could be readily visualized with Coomassie, no other proteins appeared to co-purify in stoichiometric amounts that might be able to account for the observed molecular mass of MBD1 (Figure 3.7A). Even when extracts were made at physiological ionic strength using benzonase nuclease to

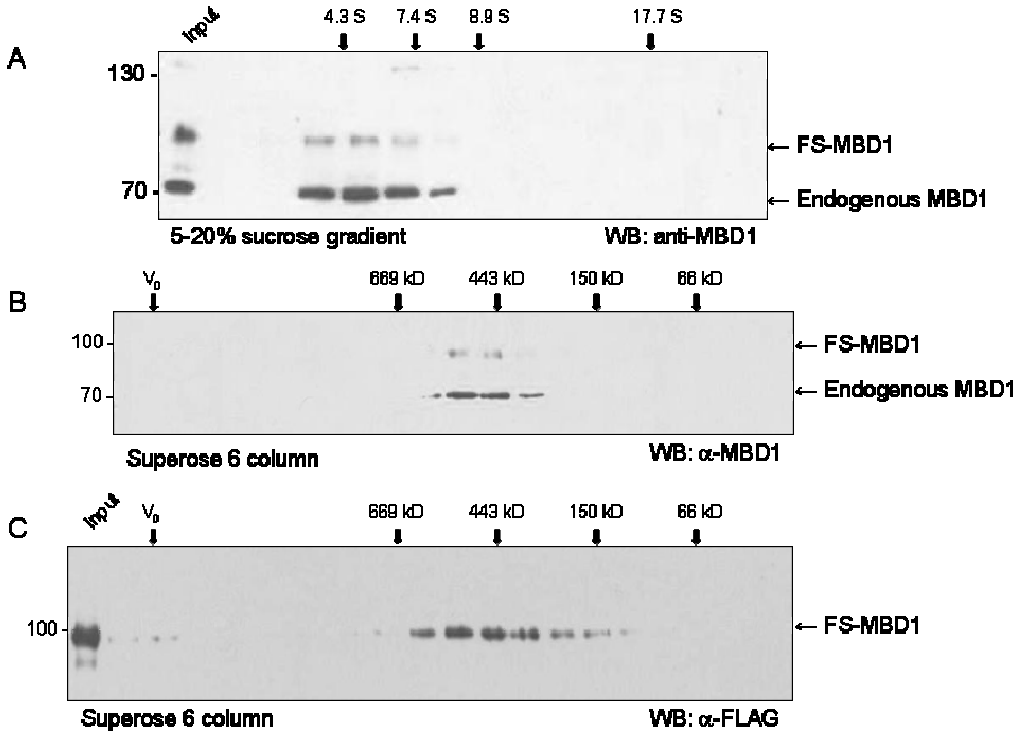

Figure 3.6 – Affinity tagging of MBD1 does not alter its Stoke's radius or S value

(A) Nuclear extracts from HeLa cells expressing FS-MBD1 were applied to a 5-20% sucrose gradient and the fractions analysed by Western blotting with MBD1 antibodies. The S values for FS-MBD1 and the endogenous protein are indistinguishable.
(B) Superose 6 size exclusion chromatography of nuclear extracts from HeLa cells expressing FS-MBD1. Fractions were analysed by Western blot with antibodies against MBD1. The Stoke's radius of FS-MBD1 is indistinguishable to that of native MBD1.
(C) As above except the gel filtration fractions were analysed by Western blot with antibodies against the FLAG epitope. This confirms that FS-MBD1 has a very similar Stoke's radius to the endogenous protein.

solubilize chromatin, only the histones were found to strongly co-purify with FS-MBD1 (Figure 3.7B). Control experiments performed in parallel using putative chromatin remodelling enzyme LSH tagged and expressed in a similar way confirmed that FS-MBD1 was indeed responsible for the major band seen on the gels. Furthermore, Western blot analysis revealed that SETDB1, a factor which has been published to interact with MBD1, is absent amongst the co-purifying proteins (Figure 3.7C).

The failure to observe MBD1 binding partners when they might have been predicted to exist has a number of plausible explanations. For instance it could be the case that any MBD1 interacting proteins were removed during the wash steps of the FS-MBD1 purification. Alternatively, it could be the case that MBD1 self associates, but that this possibility had seemed unlikely due to aberrant behaviour of bacterially expressed MBD1. In order to try to distinguish these possibilities the biophysical properties of the purified FS-MBD1 shown below (Figure 3.7A) were analysed. Surprisingly, it was not possible to detect this protein in the fractions following size exclusion analysis on a Superose 6 column (data not shown). This could be due to a combination of the dilution of the protein inherent in this technique as well as binding of purified MBD1 to the gel filtration matrix. However, sucrose gradient sedimentation revealed that some of this protein has a similar S value to the endogenous MBD1 found in HeLa nuclear extracts with a fraction of FS-MBD1 also forming an aggregate at the bottom of the gradient (Figure 3.8A). A similar investigation of FS-MBD1 purified from nuclease solubilized chromatin revealed the protein to fractionate in three peaks (marked by asterisks). Firstly some of the protein migrated in the same fashion as that in the previous experiment, a second peak further down the gradient appeared to correspond to nucleosome-associated MBD1 and finally a third peak at the bottom of the gradient suggested some aggregation of this protein too (Figure 3.8B). These observations argue that self association may account for the observed molecular mass of MBD1. Given that the derived mass calculated from the Siegel and Monty equation (136 kDa) is close to double the theoretical mass of MBD1 (61 kDa) it seems most likely that the protein forms dimers.

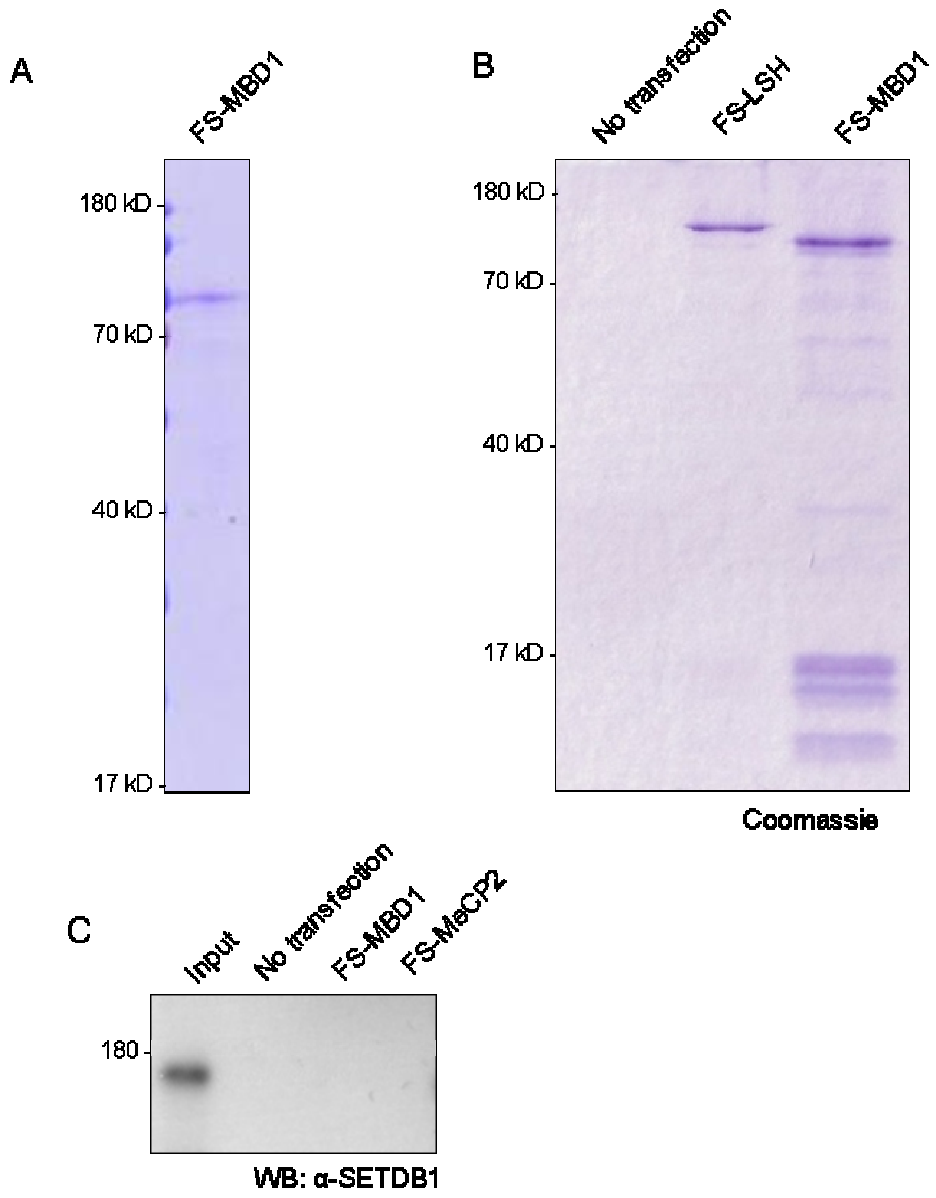

Figure 3.7 – Absence of proteins strongly co-purifying with FS-MBD1

(A) Coomassie stained gel of FS-MBD1 expressed in HeLa cells and purified from nuclear extracts using the M2 monoclonal FLAG antibody reveals no abundant co-purifying proteins.
(B) Coomassie stained gel of FS-MBD1 purified at physiological ionic strength from nuclease solubilized chromatin. Only the histones strongly co-purify.
(C) Western blot analysis using SETDB1 antibodies reveals that this protein does not co-purify with FS-MBD1 or the control protein FS-MeCP2.

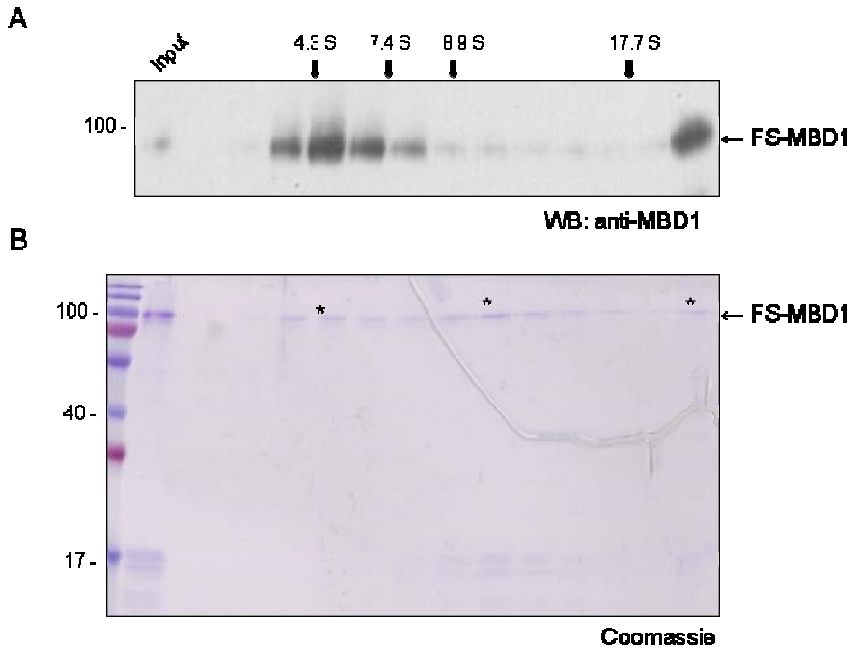

Figure 3.8 – Purified FS-MBD1 maintains its S value

(A) Western blot analysis of sucrose gradient fractions reveals that the purified MBD1 analysed in this way elutes in the same manner as does the native protein in extracts. (B) Coomassie stained gel of fractions from a sucrose gradient analysing FS-MBD1 purified at mild ionic strength from nuclease solubilized chromatin.

3.7 Co-immunoprecipitation of MBD1 over-expressed in HeLa cells

MBD1 self association was tested further by co-immunoprecipitation assays. To this end 3xFLAG-tagged MBD1 (amino acids 1-395) and untagged full length MBD1 were simultaneously over-expressed in HeLa cells. A truncated form of MBD1 was used in addition to the full length protein in order that the two could be resolved by Western blotting. Amino acids 1-395 were selected as experiments discussed in chapter six (performed by Alex Tuck) using *in vitro* translated MBD1 suggested that this portion of the protein was sufficient for self association. As a control 3xFLAG-tagged MeCP2 was co-expressed with untagged MBD1. Nuclear extracts were prepared from these cells and immunoprecipitations were performed using the M2 monoclonal antibody against the

FLAG epitope. Co-immunoprecipitation of untagged MBD1 was assessed by Western blotting using antibodies against this protein and found to occur with FS-MBD1(1-395) but not FS-MeCP2 (Figure 3.9A). Western blotting with anti-FLAG antibodies confirmed that both 3xFLAG-tagged proteins had been successfully expressed and precipitated (Figure 3.9B). The experiment described above adds weight to the suggestion that MBD1 interacts with itself. Also the fact that the transcriptional repression domain (TRD) is absent from FS-MBD1(1-395) further suggests that the interaction is mediated either by the MBD or the zinc fingers of MBD1.

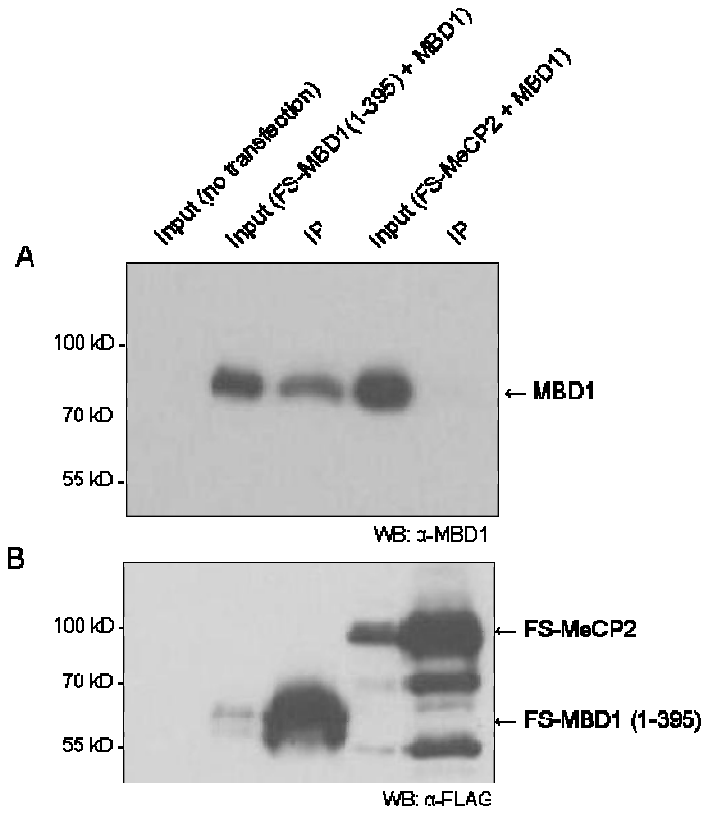

Figure 3.9 – Self association of MBD1 by co-immunoprecipitation

(A) FS-MBD1 (1-395) or FS-MeCP2 were expressed in HeLa cells with full length untagged MBD1. Nuclear extracts were subject to immunoprecipitation with an antibody against the FLAG epitope and bound proteins were analysed by Western blotting with MBD1 antibodies. An interaction between MBD1 and itself but not MeCP2 is detected. (B) Western analysis with antibodies against the FLAG epitope reveals both FS-MBD1 (1-395) and FS-MeCP2 were successfully expressed and immunoprecipitated.

4. Chapter Four - Proteomics of MBD protein partners

4.1 Nuclease allows extraction of MBD1 at mild ionic strength

As shown in chapter three, MBD1 appears to be devoid of stably and stoichiometrically associated binding partners in HeLa nuclear extracts. This is surprising given that an earlier report found MBD1 to be stoichiometrically associated with SETDB1 in these cells (Sarraf and Stancheva, 2004). The absence of stable binding partners has also been reported for MeCP2 as well as the putative chromatin remodelling enzyme LSH (Klose and Bird, 2004; Myant and Stancheva, 2008). One possible interpretation of these observations is that these proteins function autonomously and do not interact adaptively with other proteins inside the cell. However, this seems unlikely as an interaction with DNMT1 is required for LSH to function as a transcriptional repressor in reporter assays (Myant and Stancheva, 2008) and MeCP2 appears to silence transcription via a substoichiometric association with the co-repressor Sin3a (Nan *et al*, 1998).

The ionic strength of the buffers used to extract nuclear proteins may have contributed to the failure to find stable binding partners of proteins such as MBD1, MeCP2 and LSH. In order to try to exclude or overcome this problem, attempts were made to extract MBD1 from the nucleus under milder conditions. This was achieved by using the promiscuous nuclease benzonase to solubilize chromatin and its associated proteins. Western analysis of the extracts made using various salt concentrations after benzonase or control treatment of HeLa nuclei revealed that nuclease treatment caused more MBD1 to be extracted at mild ionic strength (Figure 4.1A). Examination of the same extracts by SDS-PAGE followed by Coomassie staining showed that benzonase treatment also increased the total amount of protein released from isolated nuclei at physiological ionic strength (Figure 4.1B). Also MBD1 from mouse embryonic fibroblasts appeared to be more tightly associated with chromatin than the protein in HeLa cells, and the increase in extraction efficiency when benzonase is employed was even greater (Figure 4.1C).

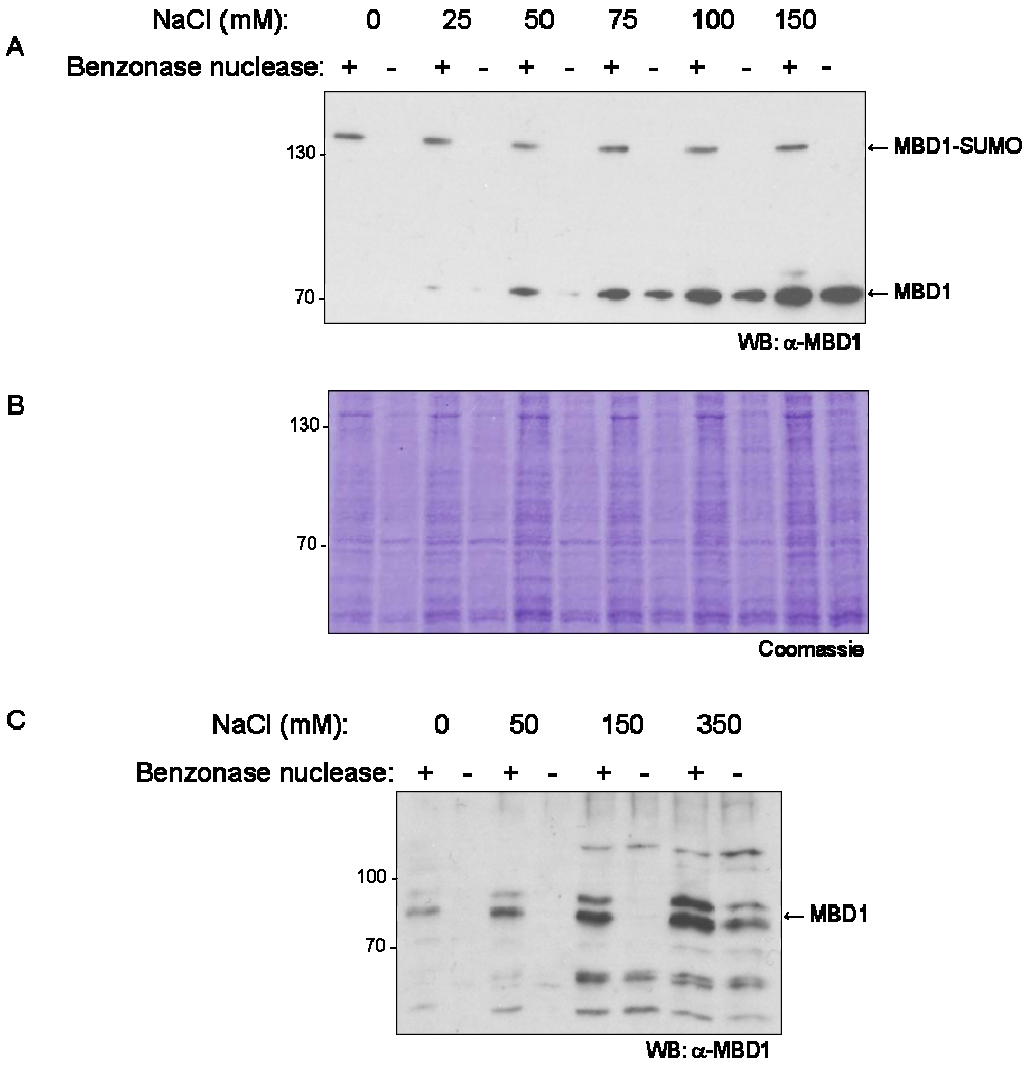

Figure 4.1 – Extraction of MBD1 from nuclei treated with benzonase

(A) Western blotting with antibodies against human MBD1 reveals more of this protein is extracted from HeLa nuclei across a range of sodium chloride concentrations after pre-treatment with benzonase.
(B) Analysis of the same extracts by SDS-PAGE followed by Coomassie staining reveals that benzonase treatment leads to an increase in the total protein extracted from HeLa nuclei at a range of sodium chloride concentrations.
(C) Western blotting with antibodies against mouse MBD1 reveals more MBD1 is extracted from MEF nuclei at different salt concentrations after benzonase treatment. Mouse MBD1 appears to be more resistant to salt extraction than the human protein.

An earlier report showed that MeCP2 can be solubilized by treatment of nuclei with micrococcal nuclease (Meehan *et al*, 1992) and so this protein seems to behave similarly to MBD1. The dual specificity of benzonase for DNA and RNA has the additional advantage that complexes purified from extracts made using this enzyme should not be contaminated by long-range indirect RNA-mediated interactions between basic proteins. This concern is particularly relevant for the MBD-containing proteins under examination here which have all been reported to interact strongly with RNA *in vitro* (Jeffery and Nakielny, 2004).

4.2 Purification of 3xFLAG-tagged chromatin-associated proteins

With the extraction procedure described above in hand, attempts were made to uncover binding partners of MBD1, MeCP2 and LSH that may previously have been missed due to the use of extraction buffers of higher ionic strength. To this end these proteins were transiently expressed in HeLa cells as fusions with N-terminal 3xFLAG and S tags. This combination of tags was chosen because, as demonstrated in chapter three, the Stoke's radius and sedimentation coefficient of exogenous MBD1 tagged in this fashion are indistinguishable from the values for the endogenous protein. Also the Stoke's radius of MeCP2 expressed with these tags in HeLa cells appeared to be the same as that found for the native protein in rat brain by Klose and Bird (data not shown). Mbd2 bearing a 3xFLAG tag was employed as a positive control because the NuRD complex would be predicted to readily co-purify with this protein. Furthermore, comparing four different proteins in parallel has the advantage of guarding against incorrectly designating factors that bind to the affinity resin or tag as interacting with the protein of interest. Such events should be recognised by the fact that they occur in all purifications carried out. It is especially important to control for this in the experiments described here due to the mild conditions used. Also the proteins under scrutiny are expected to bind to the nucleosomal DNA released by benzonase, and therefore co-purification is possible due to residence on the same piece of DNA rather than a genuine direct protein-protein interaction. Some of these artefactual identifications would be expected to occur for

multiple DNA binding proteins and so by comparing several factors it was reasoned that it should be possible, at least to some degree, to guard against being misled by this caveat.

Purification of these over-expressed proteins was accomplished by immunoprecipitation using an antibody against the 3xFLAG tag. A single step purification was used so that the duration of the experiment would be limited and complexes, which occur *in vivo* but gradually dissociate *in vitro*, might still be detected. Extraction, immunoprecipitation and wash buffers were used containing only 150 mM sodium chloride in order to try to preserve protein-protein interactions. Bound proteins were removed from the affinity matrix by competitive elution using the 3xFLAG peptide. Elution of the isolated proteins in this way allows their analysis by SDS-PAGE and staining with Coomassie or silver without the risk of a binding partner being missed due to co-migration with the antibody chains in the gel. A mock purification from cells that had not been transfected yielded no proteins that could be visualized by Coomassie staining (Figure 4.2A – lane 1) and the same procedure using cells transiently expressing tagged LSH gave a largely homogenous preparation of this protein (Figure 4.2A – lane 2). Taken together these observations support the view that the conditions being used in this experiment were not so mild as to lead to a high signal to noise ratio. The MBD1 sample contained more substoichiometric co-purifying proteins as assessed by Coomassie staining and these potentially represent binding partners. This sample also contained a large amount of histones indicating that MBD1 associates with nucleosomes under these conditions (Figure 4.2A – lane 3). The purified MeCP2 sample was also found to contain a large amount of histones as well as two major degradation products (marked by asterisks) as revealed by Western blotting with antibodies against the 3xFLAG tag (Figure 4.2B) and a variety of faintly staining bands only just visible overleaf (Figure 4.2A – lane 4). In contrast the parallel experiment with MBD2 yielded many bands that stained strongly with Coomassie some of which were of greater molecular weight than MBD2 and therefore could not simply correspond to degradation products (Figure 4.2A – lane 5).

By and large, as would be expected, the proteins in this sample correspond to the well characterized MBD2-containing NuRD complex (see sections 4.3 and 4.4).

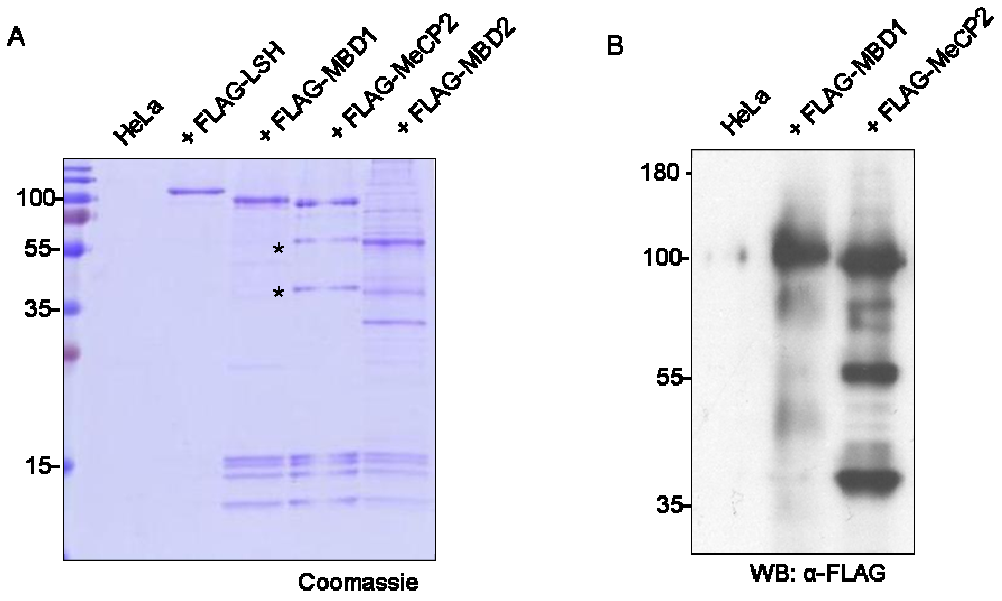

Figure 4.2 – Purification of 3xFLAG-tagged DNA binding proteins from nuclease solubilized chromatin using the M2 monoclonal antibody

(A) 3xFLAG tagged LSH, MBD1, MeCP2 and MBD2 were transiently expressed in HeLa cells, precipitated using the M2 monoclonal antibody and eluted using the 3xFLAG peptide. Proteins were then separated by SDS-PAGE and visualized by staining with Coomassie.
(B) Western blotting with antibodies against the 3xFLAG tag reveals that the major bands other than the full length protein in the lane with MeCP2 correspond to degradation products.

4.3 Identification of co-purifying proteins by mass spectrometry

In some respects the experiment outlined above could be considered as a failure because none of LSH, MBD1 or MeCP2 co-purified with striking amounts of non-histone proteins as assessed by SDS-PAGE and staining with Coomassie. However, by using mass spectrometry, it is possible to identify even low abundance proteins in relatively complex mixtures. Therefore, in an attempt to identify novel substoichiometric binding partners for the proteins under investigation, samples were prepared in the same way and then analysed by mass spectrometry in collaboration with the laboratory of Dr Juri Rappsilber. Eluting proteins from the affinity matrix using the 3xFLAG peptide rather than by denaturation facilitates the identification of low abundance peptides as the mass spectra are not complicated by the presence of a large excess of antibody. For a similar reason, prior to analysis by mass spectrometry, the samples were run a short distance in an 18% acrylamide gel in order to separate the abundant histones from the rest of the mixture. This also has the advantage of removing the 3xFLAG peptide from the preparations to be analysed.

Complete lists of the proteins identified by mass spectrometry as co-purifying with each of LSH, MBD1, MeCP2 and MBD2 are given in appendix II. In summary 150 proteins were identified in the sample generated from the mock purification from untransfected cells, 159 proteins were found with LSH, 183 proteins with MBD1, 177 with MeCP2 and 188 with MBD2. Inspection of the list of protein co-purifying with MBD2 shows that this technique allowed successful enrichment for the NuRD complex. Of the eighteen proteins reported to be associated with this complex (Le Guezennec *et al* 2006) twelve were identified in the experiment described here and in all cases on the basis of more peptides than found with any of the other bait proteins used. Also PRMT1 was identified in this sample consistent with another report on an interaction with this protein and arginine methylation of MBD2 (Tan and Nakielny, 2006). It can therefore be concluded that the method employed here is able to accurately isolate and identify genuine protein-protein interactions.

The lists of proteins identified were then filtered to remove entries that are likely to represent contaminants rather than biologically relevant binding partners. Proteins found to co-purify with two or more baits were discarded as were those identified on the basis of only one or two peptides. Finally proteins annotated as 'mitochondrial' were removed as were the bait proteins, when identified in the sample analysed subsequently in the mass spectrometer. The filtered lists of proteins co-purifying with LSH, MBD1 and MeCP2 are given below (Table 1). The unique proteins found with MBD2 were not considered further at this stage because the stringent filtering criteria employed led to the subtraction of many components of the NuRD complex due to their low level co-purification with other proteins examined (see appendix II and section 4.4).

Few of the proteins listed with LSH are likely to represent genuine interaction partners. Most of them are either not localized to the nucleus or are found in complexes represented by different subunits as co-purifying with one of the other baits tested here. In some regards this is similar to what is observed in the list of proteins identified in the mock purification from untransfected cells. Furthermore, the DNA methyltransferase enzymes, which associate with LSH (Zhu *et al*, 2006; Myant and Stancheva, 2008), were not found even in the unfiltered list of proteins co-purifying with this factor. However, the situation with the samples containing MBD1 and MeCP2 was more promising. Among the 33 polypeptides identified uniquely in the MeCP2 experiment lie ATRX, two components of the Sin3 co-repressor complex (Sin3a and SAP180), four components of the nuclear co-repressor complex (NCoR1, HDAC3, TBL1X and TBL1XR1) and BAF60a which is associated with the BRG1 chromatin remodelling complex. All of these complexes are documented as interacting with MeCP2 (Nan *et al*, 2007; Nan *et al*, 1998; Kokura *et al*, 2001; Harikrishnan *et al*, 2005). Likewise in the case of MBD1 both RING2 and CBX8 were both among the 26 proteins identified as present in the sample. These two proteins are components of a polycomb repressive complex 1 (Bárdos *et al*, 2000) and such a complex has been reported to interact with MBD1 (Sakamoto *et al*, 2007).

As well as the rediscovery of several known interactions, novel candidate binding partners for both MBD1 and MeCP2 were also uncovered. Among the factors found uniquely with MBD1 were the histone H3 lysine K4 demethylase LSD1, the histone acetyltransferase MYST2, the centromere component CENP-B, and two homologues (SUHW3 and SUHW4) of the Drosophila insulator protein suppressor of hairy wing (Dorsett, 1993). LSD1 has been implicated in transcriptional repression (Shi *et al*, 2004; Lee *et al*, 2005) as have the yeast homologues of MYST2 (Reifsnyder *et al*, 1996). Also MBD1 is enriched at centromeres on metaphase chromosomes (Ng *et al*, 2000) and DNA methylation is known to regulate the function of another insulator protein, CTCF (Hark *et al*, 2000). Therefore all of these represent biologically plausible interactions which merit further investigation. In the case of MeCP2 the polymerase associated factor (PAF) complex was identified by the presence of four subunits (Parafibromin, PD2, Ctr9 and WDR61) all of which are absent in all the other samples. This potential interaction with the PAF complex, a group of proteins involved in transcriptional elongation (Chaudhary *et al*, 2007), could imply a hitherto unsuspected role for MeCP2 in the regulation of this process. MeCP2 also pulls down the transcription elongation factor SPT6 and its interaction partner IWS1 (Brès *et al*, 2008) augmenting the temptation to speculate that this protein plays a role downstream of the initiation of transcription.

Some of the other proteins identified should be treated with extra caution as some are components of well characterized complexes represented here by other subunits which co-purify with different baits. For example, the histone H3 lysine K9 methyltransferase G9a is identified as co-purifying with MBD1 alone but GLP, a similar enzyme which is reported to make stable heterodimers with G9a (Tachibana *et al*, 2005) is found in both the MBD1 and MeCP2 containing samples. Likewise EED, a component of the polycomb repressive complex 2, is found only with MBD1 whereas EZH2, the catalytic component of this complex, is again found with both MBD1 and MeCP2. Similarly, although the chromatin assembly factor ACF1 was found uniquely with MeCP2, its binding partner ISWI (Bochar *et al*, 2000) was found to co-purify with all three MBD-containing proteins.

LSH (88)	MBD1 (32)	MeCP2 (43)
Keratin, type I cytoskeletal 9 (25)	TIF1-alpha (12)	Parafibromin (16)
AHNAK (8)	RNAPII large subunit (10)	Nipped-B-like protein (14)
Junction plakoglobin (6)	SUHW4 (8)	SPT6 (13)
26S proteasome non-ATPase regulatory subunit 1 (5)	FLYWCH-type zinc finger 1 (8)	TBL1XR1 (13)
Annexin A2 (5)	CKII subunit alpha (7)	NCoR1 (13)
Elongation factor 1-delta (4)	G9a (7)	PD2 (11)
hnRNP K (4)	TFIIIC1 (7)	Ctr9 (11)
Long-chain-fatty-acid-CoA ligase 3 (4)	BPTF (6)	ligase III, DNA, ATP-dependent (10)
DDOST (3)	EED (6)	Spectrin beta chain, brain 1 (9)
4F2 cell-surface antigen (3)	LSD1 (6)	Probable RNA helicase DDX46 (9)
HSP90 beta (3)	MCM4 (5)	ACF1 (9)
WDR76 (3)	HLTF (5)	40S ribosomal protein SA (8)
NS1A binding protein (3)	SPT5 (5)	Sin3a (8)
U4/U6.U5 tri-snRNP-associated protein 1 (3)	MAX dimerization protein 5 (5)	SAP180 (7)
Spindlin-1 (3)	CENPB (4)	Keratin, type II cytoskeletal 6E (7)
No Transfection	TRIM26 (4)	Importin alpha-3 subunit (6)
Desmoplakin (9)	MYST2 (4)	APC subunit 1 (6)
JUP protein (8)	CBX8 (3)	IWS1 homolog (6)
Keratin 5b (6)	Cartilage-associated protein (3)	Bifunctional polynucleotide phosphatase/kinase (6)
Splicing factor 3 subunit 1 (4)	RING2 (3)	P37 AUF1 (5)
Similar to E2F7 (4)	β-catenin-like protein 1 (3)	ANP32B (5)
Keratinocyte proline-rich protein (4)	PHD interacting protein (3)	DNMT3B (5)
Transglutaminase 3 E (3)	Prolyl 3-hydroxylase 2 (3)	APC subunit 3 (5)
Keratin, type II cytoskeletal 2 oral (3)	SUHW3 (3)	HDAC3 (4)
KIAA1546 protein (Fragment) (3)	Zinc finger protein 262 (3)	WDR61 (4)
Pre-mRNA-processing factor 19 (3)	Zinc finger protein 592 (3)	H3.3 (4)
Paraspeckle protein 1 (3)		MCM5 (4)
NCBP1 (3)		TBL1X (4)
		ANP32A (3)
		BAF60a (3)
		ATRX (3)
		similar to Snf2-related CBP activator protein (3)
		hnRNP Q (3)

Table 1 – Lists of factors co-purifying with LSH and MBD-containing proteins

Only proteins represented by three or more peptides and identified uniquely with one of the baits are listed. Number of peptides observed given in brackets. Components of complexes which represent known interactions are shown in red.

4.4 Validation of mass spectrometry data by Western blotting

The next step in this line of research was to investigate which of the candidates identified represent genuine interaction partners of the proteins under investigation. To this end the experiment described above was repeated except co-purifying proteins were analysed by SDS-PAGE followed by Western blotting with various antibodies rather than mass spectrometry. This has the benefits of being somewhat more quantitative, establishing reproducibility and verifying the presence or absence of co-purifying proteins by an independent technique.

Probing with antibodies against HDAC1 revealed a pattern that might have been predicted from the mass spectrometry data and what is reported in the literature. The protein is absent in the mock purification using untransfected cells, a weak signal is present in the LSH sample with a slightly stronger signal in the MBD1 and MeCP2 samples and a stronger signal still in the lane containing MBD2 (Figure 4.3 – top panel). Using antibodies against EZH2 and GLP also suggested reproducibility as consistent with the mass spectrometry data both of these proteins were found most strongly with MBD1 and MeCP2 (Figure 4.3 – second and fourth panels). G9a is also apparently present predominantly in both of these samples (Figure 4.3 – fifth panel) in spite of being unique to MBD1 as assessed by mass spectrometry, although as discussed in the previous section, this is not a great surprise due to the presence of GLP with both MBD1 and MeCP2. Western blot analysis of LSD1 revealed that it too co-purified with both MBD1 and MeCP2 (Figure 4.3 – third panel). This is slightly surprising given that LSD1 was only found with MBD1 by mass spectrometry. The reason for this discrepancy is unclear and so this emphasises the need to confirm the candidates found by mass spectrometry using an alternative technique to analyse biologically independent purifications. Finally, Western blotting with antibodies against HDAC3 and ACF1 bears out exactly what was found in the previous sections with both of these proteins co-purifying uniquely with MeCP2 (Figure 4.3 – sixth and seventh panels). Given that ISWI, a stable binding partner of ACF1, was found by mass spectrometry to co-purify

with MBD1 and MBD2 as well as MeCP2, there was a possibility that ACF1 would not be found uniquely with MeCP2. Therefore, even the proteins listed in section 4.3 that are viewed with caution for this kind of reason can still merit further investigation.

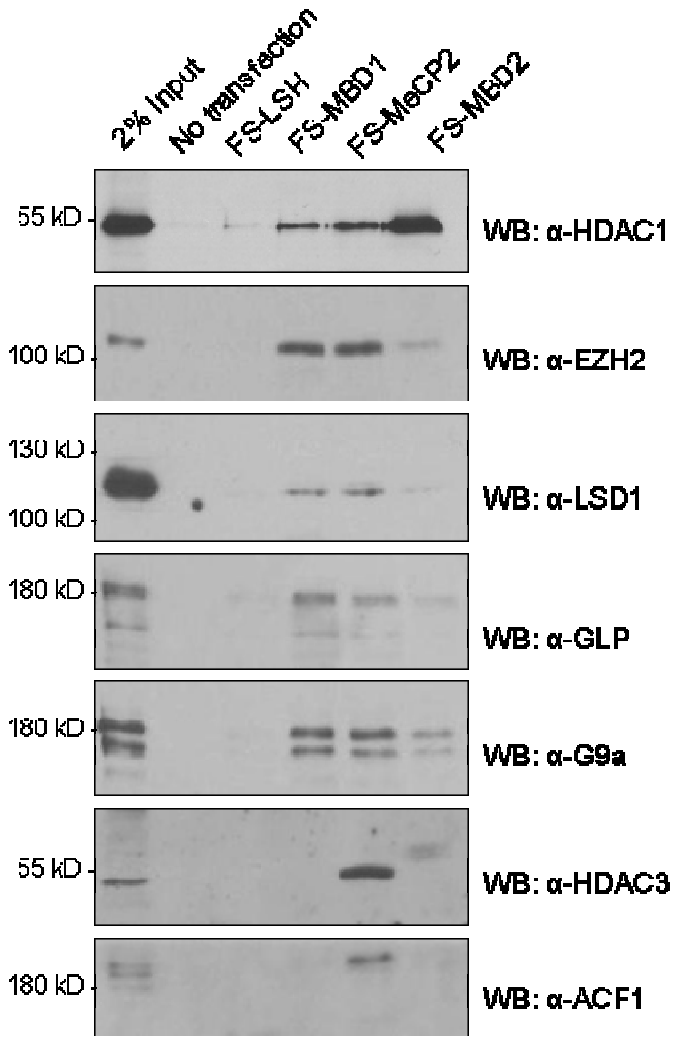

Figure 4.3 – Western blots for factors co-purifying with LSH and MBD proteins

Purifications were carried out as for the mass spectrometry experiment. Samples were then analysed by Western blotting. Most proteins are present in the predicted samples based on the mass spectrometry data. LSD1 is unexpectedly found in the MeCP2 sample. It is conformed that HDAC3 and ACF1 co-purify uniquely with MeCP2.

4.5 Histone modifications associated with MBD proteins

Post translational modifications of the histones are believed to regulate many aspects of chromatin structure and function. Such modifications are known to recruit proteins to specific regions of the genome and thus regulate processes such as DNA repair (Sanders *et al*, 2004) and DNA methylation (Ooi *et al*, 2007). It is therefore possible that in addition to DNA sequence and CpG methylation, histone modifications might be involved in targeting different MBD proteins to different regions of the genome. This is a plausible scenario for MBD2 as the p66α and p66β components of the NuRD complex both interact preferentially with non-acetylated histone tails *in vitro* (Brackertz *et al*, 2006). Also, recombinant nucleosome-bound MeCP2 can be chemically cross-linked to histone H3 *in vitro* (Nikitina *et al*, 2007) indicating a close contact between these two proteins with the potential to be regulated by a modification.

As well as the possibility that histone modifications recruit MBD-containing proteins, it has also been reported that these proteins in fact recruit enzymes which covalently modify histones. MBD2 is a component of a histone deacetylase nucleosome remodelling (NuRD) complex and MeCP2 associates with the histone deacetylase containing Sin3a co-repressor complex as well as a previously unidentified histone H3 lysine K9 methyltransferase activity (Fuks *et al*, 2003). Furthermore, MBD1 is reported to interact with histone deacetylases as well as H3-K9 methyltransferase activities (Villa *et al*, 2006; Fujita *et al*, 2003; Sarraf and Stancheva, 2004). Therefore it is of interest to investigate whether specific histone modifications might be found on the nucleosomes associated with different MBD-containing proteins. Given that nucleosomes strongly co-purify with MBD1, MeCP2 and MBD2, it is possible to ask by mass spectrometry whether specific post-translational modifications are found on these histones. Therefore, in collaboration with the laboratory of Dr Axel Imhof, the bands corresponding to histones H3 and H4 co-purifying with each of MBD1, MeCP2 and MBD2 were cut out a Coomassie stained gel similar to the one shown above (Figure 4.2A) and analysed by

mass spectrometry. As a control, the H3 and H4 bands from the input lane from an extract used for a mock purification were cut out and similarly analysed.

The mass spectra generated provide information on the relative abundance of only a subset of potential histone modifications. The peaks corresponding to unmodified and acetylated histone H4 amino acids 4-17 did not vary strikingly between samples (data not shown). Likewise the relative amounts of the peptides corresponding to histone H3 residues 9-17 in the monomethylated, dimethylated, trimethylated/acetylated or unmodified state did not vary substantially between preparations (data not shown). A zoom of the region of the spectrum corresponding to various states of histone H3 amino acids 27-40 is displayed blow (Figure 4.4). It is difficult to discern alterations in the relative levels of unmodified, trimethylated/acetylated, dimethylated or $K27Me_2K36Me_3$ peptides, however, the peptide corresponding to $K27Me_2K36Me_2$ is almost two-fold enriched in the MBD2 sample.

The significance of this observation is somewhat doubtful as the data presented only represent a single experiment. To pursue this line of enquiry it would therefore be necessary to validate the result presented here by Western blot analysis of the histones co-purifying with MBD2 and other proteins. This would also enable the extension of this line of experiments to include observations of modification states not interrogated by the mass spectrometry detailed here. If the finding described above proves to be robust then it might prove fruitful to investigate whether the presence of MBD2 is a cause or a consequence of this combination of histone marks. In the latter case it would be of interest to ask which component of the NuRD complex is responsible for this targeting.

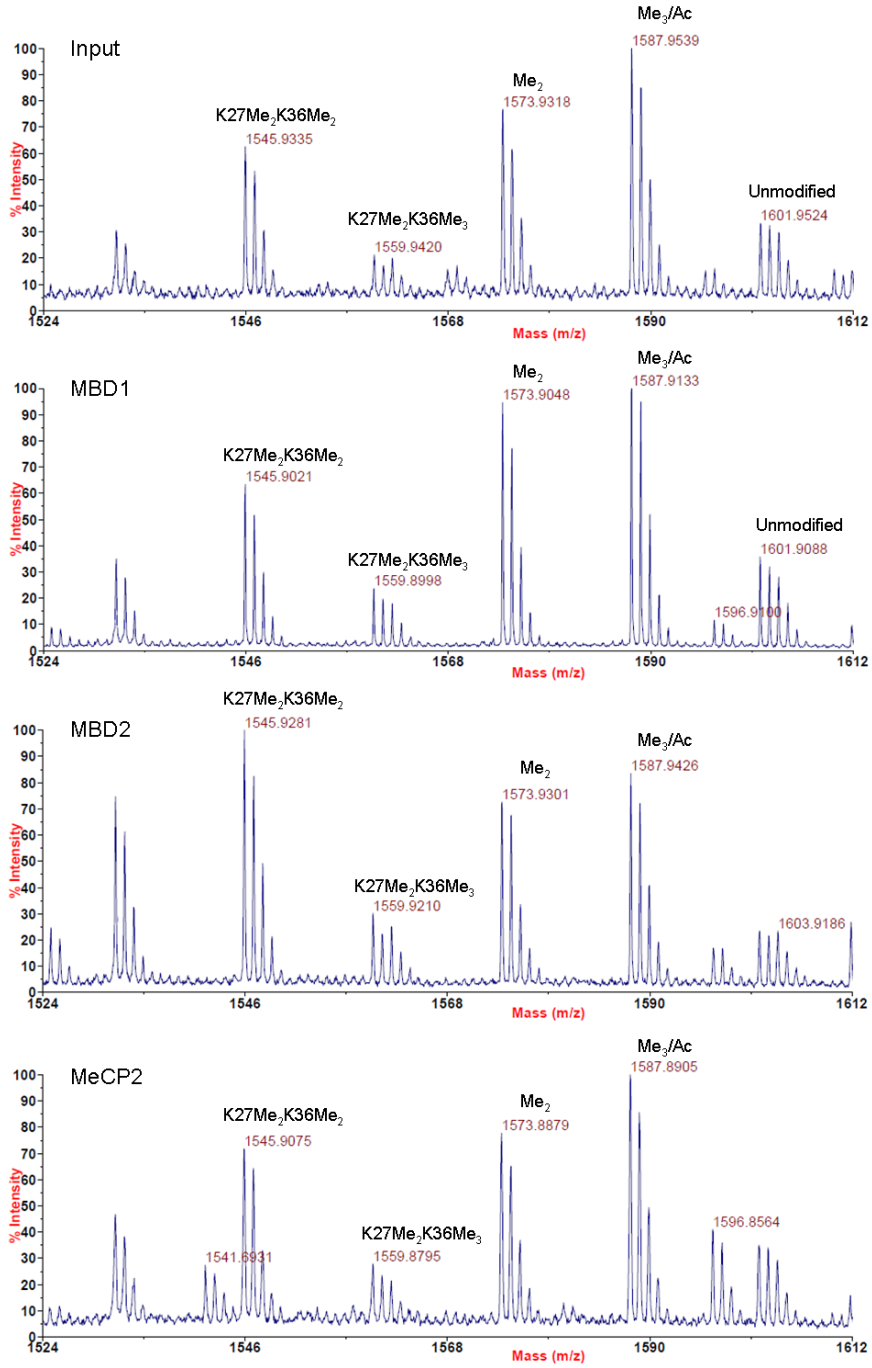

Figure 4.4 – Mass spectra showing peptides corresponding to different modification states of amino acids 27-40 of histone H3.

The peak which corresponds to the $K27Me_2K36Me_2$ appears to be enriched in the MBD2 sample relative to the spectra obtained from the input, MBD1 and MeCP2 samples. No substantial differences are observed in other modification states.

5. Chapter Five - Regulation of SUMO modification of MBD1

5.1 Ubiquitin-like modifiers

The C-termini of ubiquitin and ubiquitin like proteins can be conjugated to lysine residues in target proteins by the formation of isopeptide bonds. In addition to ubiquitin this group of proteins includes the poorly characterized NEDD8 and ISG15 as well as the SUMO (small ubiquitin-like modifier) family. These post-translational modifications increase the complexity of eukaryotic proteomes and provide a mechanism with the potential to regulate the function of almost any protein. The mammalian SUMO family comprises three different proteins; namely SUMO-1, SUMO-2 and SUMO-3. SUMO-1 shows approximately 50% sequence identity to SUMO-2 and SUMO-3 with the latter two of these proteins, which differ in only three N-terminal amino acids, yet to be functionally distinguished (Hay, 2005). The mechanism of conjugation initiates with the adenylation of the C-terminus of SUMO by the E1 activating enzyme which consists of a SAE1 and SAE2 heterodimer. The linkage with AMP is then broken to allow the C-terminal carboxyl group of SUMO to link to a sulphydryl group in SAE2. Subsequently, SUMO is transesterified to a sulphydryl in the E2 conjugating enzyme UBC9 from where it can be directly transferred to target proteins. This contrasts to the situation with ubiquitin where there is also an absolute requirement for an E3 ligase enzyme (Hay, 2005). Finally SUMO can be cleaved from substrates by specific SUMO protease enzymes and this completes the 'SUMO cycle' which is shown below (Figure 5.1).

SUMO modification was first described in mammalian cells as a covalent modification of the GTPase activating protein RanGAP1 (Matunis *et al*, 1996), which is required for proper localization of this protein to the nuclear pore complex (Mahajan *et al*, 1997). Since then many other substrates for SUMO modification have been found by proteomic analyses (Li *et al*, 2004; Vassileva and Matunis, 2004; Vertegaal *et al*, 2004; Zhao *et al*, 2004) suggesting roles for SUMO in the regulation of diverse cellular processes. This is

consistent with other studies linking SUMO to proper cell cycle progression (Johnson and Blobel, 1997; Seufert *et al*, 1995), the DNA damage response (Hoege *et al*, 2002), subcellular transport (Lin *et al*, 2003), antagonization of ubiquitination (Desterro *et al*, 1998) and the regulation of transcription (see below).

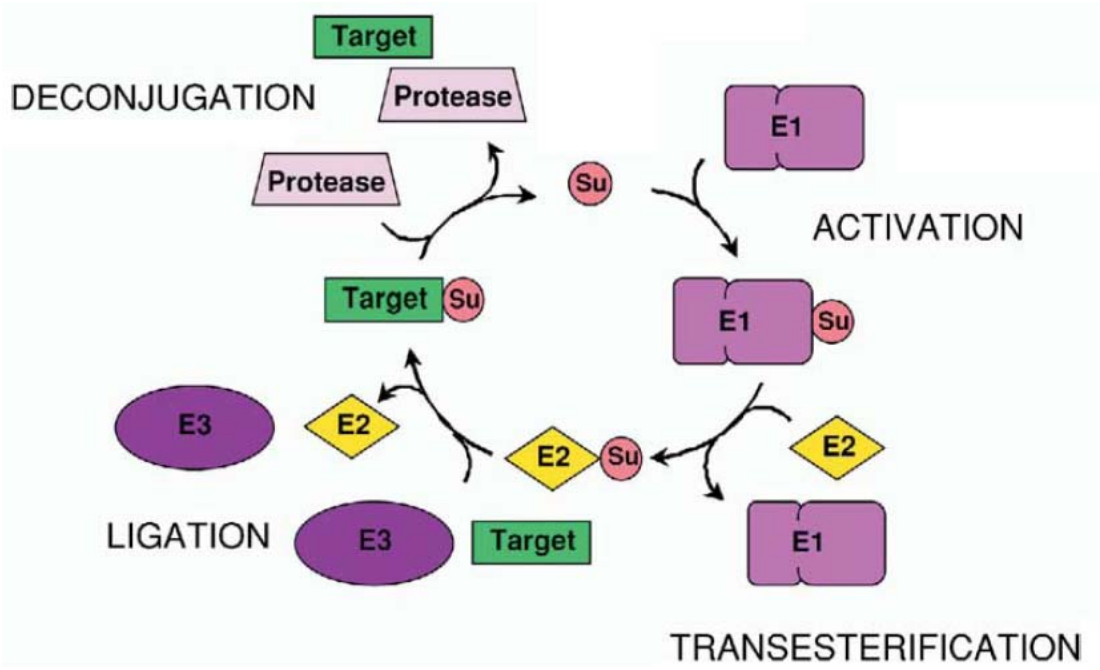

Figure 5.1 – The SUMO cycle (adapted from Hay, 2005)

SUMO is covalently joined to an E1 enzyme. It is then transferred to an E2 enzyme via a transesterification reaction before being ligated to a lysine of a substrate. The cycle is completed by the removal of SUMO from a substrate by SUMO protease enzymes.

SUMO modification of transcription factors is frequently associated with transcriptional repression (Gill, 2005). For example, there is evidence that SUMO modification of the Ets family transcription factor Elk-1 enhances its association with HDAC2 leading to transcriptional repression (Yang and Sharrocks, 2004). Furthermore, when targeted to a promoter by fusion with the Gal4 DNA binding domain, both SUMO-1 and SUMO-2 were able to silence transcription indicating that SUMO has an intrinsic repressive ability (Yang *et al*, 2003; Holmstrom *et al*, 2003). However, there are also cases where it can lead to gene activation. This appears to be so with SUMO modification of Ikaros, heat shock transcription factors, the nuclear factor of activated T cells (NFAT) and the tumour suppressor p53, under some circumstances (Gómez-del Arco *et al*, 2005; Goodson *et al*, 2001; Hong *et al*, 2001; Terui *et al*, 2004; Rodriguez *et al*, 1999).

The level of SUMO modification of individual proteins can be controlled by E3 ligase enzymes as well as SUMO protease activities, which can cleave SUMO from conjugated proteins. Although not absolutely required for modification, various SUMO E3 ligases, which can increase the efficiency of conjugation to SUMO of different substrates, have been uncovered. To date these comprise the PIAS family, RanBP2 and the Polycomb protein Pc2. Evidence exists that each enzyme promotes the modification of a different set of substrates (Melchior *et al*, 2003). For instance, over-expression of the SUMO E3 ligase PIAS1 (but not the other enzymes PIAS3, PIAS3L, PIASxα, PIASxβ, PIASγ or Pc2) led to increased SUMO modification and inhibited activity of the progesterone receptor PR-A (Jones *et al*, 2006). The first activity shown to be able to remove SUMO from substrates was Ulp1p from budding yeast (Li and Hochstrasser, 1999) and subsequently mammalian activities have been uncovered on the basis of sequence homology. Evidence that different proteases can target different SUMO modified proteins comes from budding yeast (Li and Hochstrasser, 2003). The accumulation of SUMO conjugates seen in Ulp2 deletion strains can be rescued by removal of the non-catalytic N-terminus of Ulp1. The fact that this truncation of Ulp1 leads to increased conjugation in wild-type strains suggests that the two yeast SUMO proteases have distinct sets of substrates.

MBD1 is subject to modification with SUMO-1 at two lysine residues (450 and 489) in its C-terminus. This modification indirectly disrupts the interaction of MBD1 with the co-repressor SETDB1 thus antagonizing its ability to repress transcription. Furthermore, the levels of the E3 SUMO ligases PIAS1 and PIAS3 control the amount of SUMO modification of MBD1 (Lyst *et al*, 2006). An independent study found MBD1 to be modified with SUMO-1 and SUMO-2/3, and presented evidence that this enhanced the interaction of MBD1 with MCAF leading to the formation of heterochromatin at MBD1 containing foci (Uchimura *et al*, 2006). The reasons for the apparent discrepancies between these two studies are unclear although they could involve differences in the cell lines employed. It is also unknown to date whether or not MBD1 function might be regulated by SUMO protease activities. Therefore the impact that such enzymes might have on SUMO modification of MBD1 was investigated.

5.2 Recombinant MBD1-SUMO is efficiently cleaved in vitro

Methods have recently been established for the production of recombinant SUMO modified proteins in *E.coli* (Mencía and de Lorenzo, 2004; Uchimura *et al*, 2004). Indeed, sufficient quantities of protein were produced to allow the determination of the X-ray crystal structure of thymine DNA glycosylase conjugated to SUMO-1 (Baba *et al*, 2005). A similar method was previously applied to GST-tagged MBD1 in order to test the effect of SUMO modification on binding to SETDB1 *in vitro* (Lyst *et al*, 2006). Briefly, this involved expressing GST-tagged MBD1 in *E.coli* simultaneously with HIS_6-tagged SUMO-1 as well as the E1 and E2 SUMO ligase enzymes. Conjugates were then purified from the bacterial lysates using nickel affinity followed by glutathione sepharose chromatography and this SUMO modified MBD1 was then used as a substrate to test various protease activities. When incubated with the commercially available catalytic domain of the budding yeast SUMO protease, Ulp1, GST-tagged SUMO modified MBD1 was efficiently cleaved to yield the unmodified fusion protein (Figure 5.2A and Lyst *et al*, 2006). It is not surprising that a yeast enzyme should be effective against human proteins as the isolated catalytic domain of Ulp1 is known to be highly

promiscuous with its substrate specificity normally being conferred by a regulatory domain which localizes the protein to nuclear pores (Panse *et al*, 2003).

To date six mammalian proteins (SENP1-3 and SENP5-7) have been shown to possess SUMO protease activity. Of these only SENP1 and SENP2 are preferentially active against SUMO-1 rather than SUMO-2/3 conjugates with the other four showing the opposite specificity (Mukhopadhyay and Dasso, 2007; Lima and Reverter, 2008). In the HeLa cells used in the experiments described here, MBD1 is reported to be modified only with SUMO-1 and not SUMO-2/3 (Lyst *et al*, 2006). Therefore SENP1 and SENP2 appeared to be the best candidates for a role in the regulation of the activity of MBD1, and so the activity of these enzymes was tested against recombinant SUMO-1 modified MBD1. To this end, full-length GFP-tagged SENP1 and SENP2 as well as GFP alone were transiently expressed in HeLa cells and then purified from nuclear extracts by immunoprecipitation using an antibody against GFP. These proteins were then incubated with recombinant SUMO-1 modified GST-MBD1 and deconjugation was monitored by SDS-PAGE followed by Western blotting with antibodies against GST. This revealed that both GFP-SENP1 and GFP-SENP2 but not GFP alone were able to remove SUMO from GST-MBD1 (Figure 5.2B). Furthermore, this activity was also found using limiting amounts of nuclear extracts from HeLa cells expressing GFP-SENP1 or GFP-SENP2 but not those expressing GFP alone (Figure 5.2C). This second result also argues that the activity observed in the immunoprecipitated SENP proteins is intrinsic to the enzymes rather than residing in some co-purifying factor. An activity capable of cleaving recombinant SUMO-modified MBD1 was also detected in non-limiting amounts of extracts from untransfected HeLa cells. This activity was sensitive to NEM but not EDTA, PMSF or a commercially available protease inhibitor cocktail (Figure 5.2D). It could therefore potentially explain why the levels of SUMO-modified MBD1 in HeLa nuclear extracts are dependent on the presence of NEM in the buffer (Lyst *et al*, 2006) and this activity might plausibly correspond to endogenous SENP1 or SENP2.

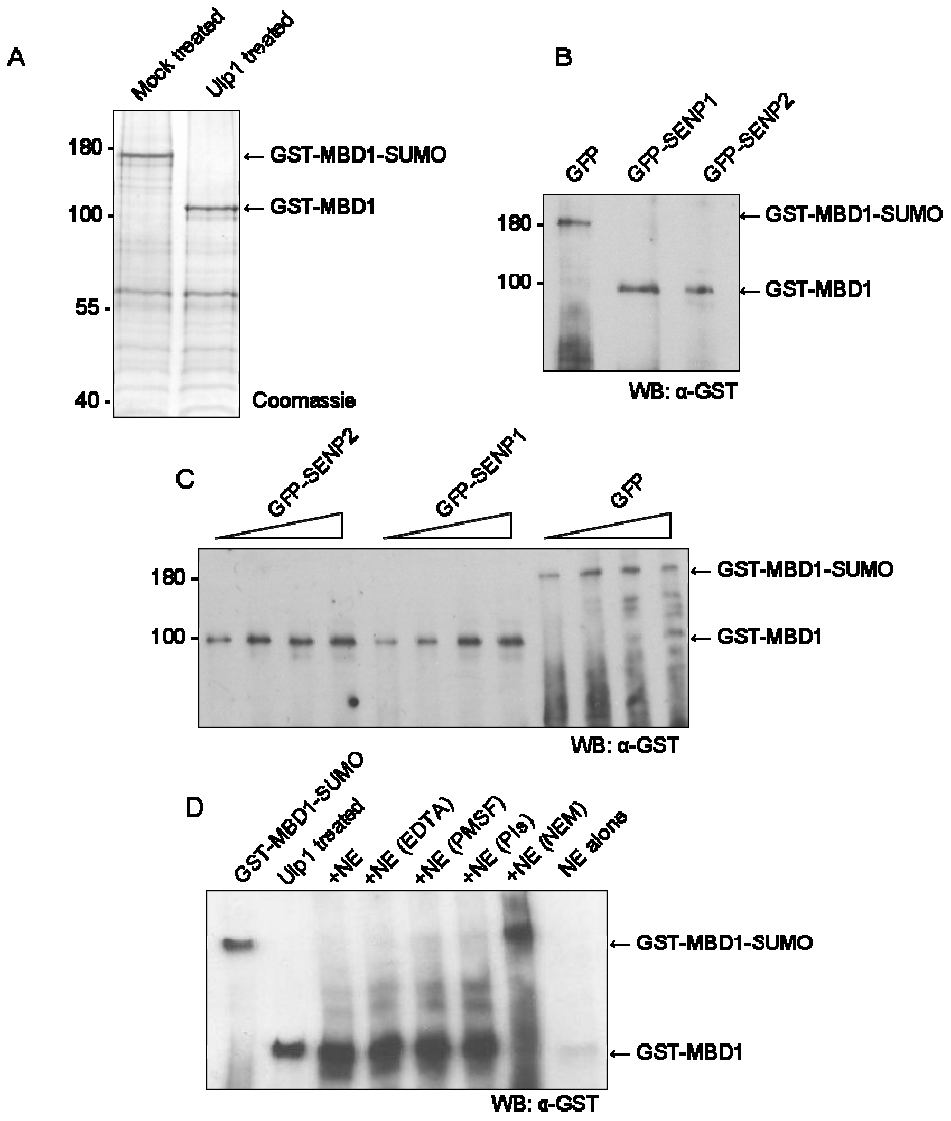

Figure 5.2 – Cleavage of recombinant SUMO modified MBD1 *in vitro*

(A) Coomassie staining of recombinant GST-tagged SUMO modified MBD1 either treated or mock treated with the SUMO protease Ulp1.
(B) Recombinant GST-tagged SUMO modified MBD1 revealed by Western blotting after incubation with either GFP, GFP-SENP1 or GFP-SENP2 that had been immunoprecipitated from HeLa nuclear extracts.
(C) Recombinant GST-tagged SUMO modified MBD1 revealed by Western blotting after incubation with limiting amounts of HeLa nuclear extracts from cells transiently expressing either GFP, GFP-SENP1 or GFP-SENP2.
(D) Recombinant GST-tagged SUMO modified MBD1 revealed by Western blotting after incubation with HeLa nuclear extracts with or without various protease inhibitors.

5.3 Native MBD1-SUMO is refractory to SUMO proteases in vivo

Given the results of the *in vitro* experiments described above, it was hypothesized that at least one of SENP1 or SENP2 might contribute to the regulation of SUMO modification of MBD1 *in vivo*. To test this notion GFP-SENP1, GFP-SENP2 and GFP alone were transiently expressed in HeLa cells, which were then directly lysed in hot Laemmli buffer. This buffer instantly denatures most proteins so that the levels of SUMO modification of MBD1 *in vivo* could be assayed without the confounding effects of SUMO ligation and removal occurring *in vitro*. Furthermore, in order to avoid the diluting effects of non-expressing cells, a small amount of plasmid DNA carrying a puromycin resistance gene was co-transfected into these cells. Having established that a concentration of 2.5 μg/ml puromycin was required to kill HeLa cells by overnight treatment, this allowed untransfected cells to be removed by overnight drug selection on the day after transfection. Fluorescence microscopy confirmed that over 95% of the surviving cells were expressing GFP or a fusion protein thereof (Figure 5.3A) and SDS-PAGE followed by Western blot analysis of extracts from these cells using antibodies against GFP revealed they were ectopically expressing fusion proteins of the predicted size (Figure 5.3B). However, probing the same membrane with antibodies against MBD1 revealed no changes in the levels of unmodified MBD1 versus SUMO modified MBD1, even when SENP1 or SENP2 were over-expressed (Figure 5.3C). It therefore seems unlikely that SUMO modification of MBD1 is regulated by these enzymes *in vivo*. The apparent inconsistency with the *in vitro* data is not a great surprise as SENP1 has previously been reported to be a much more promiscuous enzyme *in vitro* than *in vivo* (Yamaguchi *et al*, 2005).

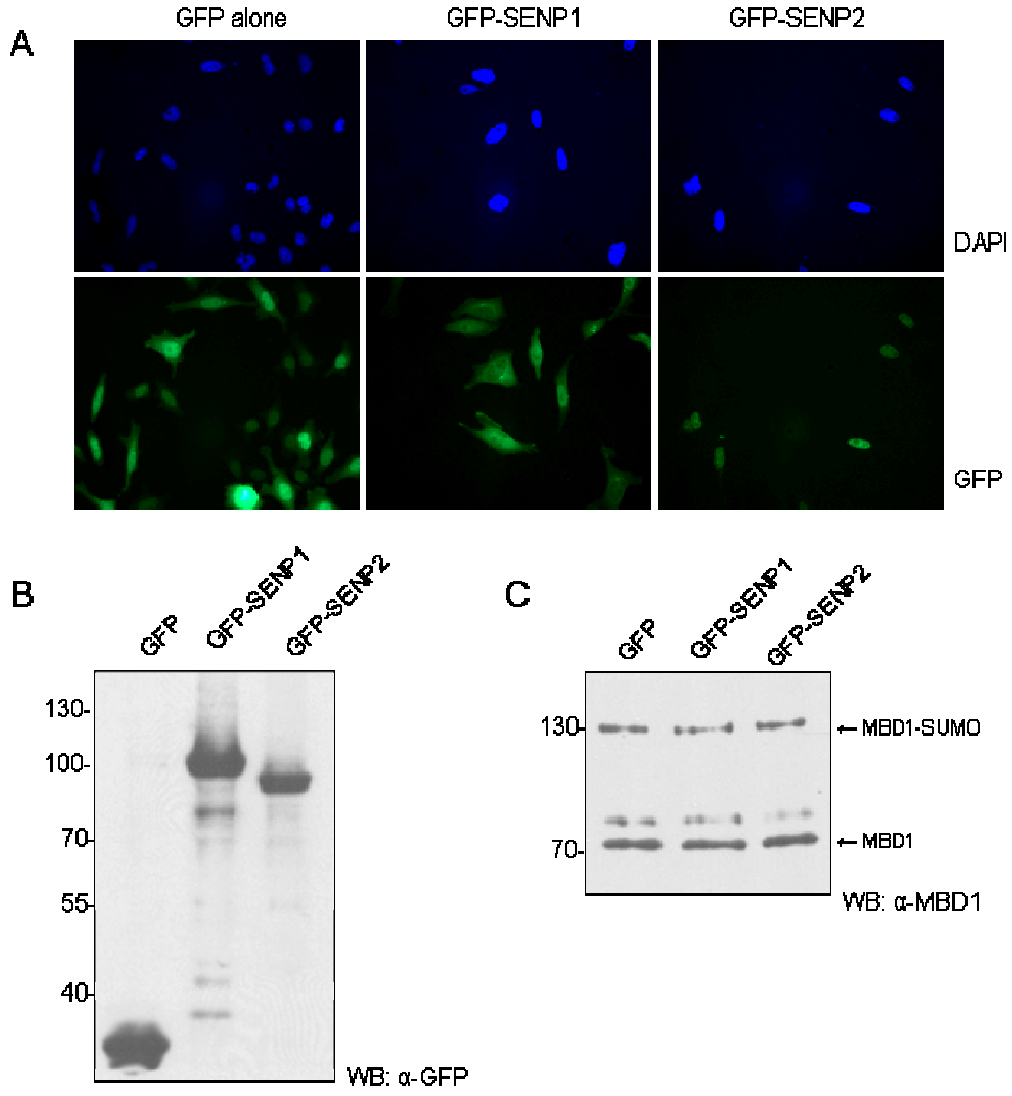

Figure 5.3 – Over-expression of SENP1 or SENP2 does not reduce levels of MBD1 SUMO modification *in vivo*

(A) Fluorescence microscopy confirms expression of GFP, GFP-SENP1 or GFP-SENP2 in over 95% of the surviving transfected cells following puromycin treatment.
(B) Western blot analysis of HeLa whole cell extracts using antibodies directed against GFP confirms expression of GFP, GFP-SENP1 or GFP-SENP2.
(C) Western blot analysis using antibodies against MBD1. Levels SUMO conjugated MBD1 and unmodified MBD1 in HeLa whole cell extracts following expression of either GFP alone, GFP-SENP1 or GFP-SENP2 are indistinguishable.

5.4 Native MBD1-SUMO is refractory to SUMO proteases in vitro

As alluded to at the end of the last section, one possible explanation for the differential susceptibility of SUMO modified MBD1 to cleavage in the experiments described above is the distinction between *in vitro* and *in vivo* analyses. Alternatively this discrepancy could arise due to the different sources of MBD1 being used as a bacterially expressed GST fusion is being compared with a native protein from HeLa cells.

In order to distinguish these possibilities the abilities of GFP-SENP1 and GFP-SENP2 to cleave SUMO from native MBD1 were assayed *in vitro*. HeLa cells were transfected with GFP-SENP1, GFP-SENP2 or GFP alone, and these proteins were purified from nuclear extracts by immunoprecipitation using an antibody against GFP. Control reactions using recombinant SUMO-modified MBD1 as a substrate confirmed that once again catalytically active SENP1 and SENP2 had been purified (Figure 5.4A). Nuclear extracts from untransfected HeLa cells containing native SUMO-modified and unmodified MBD1 were also incubated with these immunopurified SENP proteins. Western blotting with antibodies against MBD1 revealed that native SUMO-modified MBD1 was unaffected by incubation with these proteases (Figure 5.4B). Therefore the apparently contradictory results previously obtained by analysis of native MBD1 *in vivo* versus recombinant MBD1 *in vitro* are explained by differences intrinsic to the substrates rather than the reaction conditions used. This is also the case for cleavage of recombinant compared to native SUMO-modified MBD1 both by endogenous SUMO protease activity in HeLa nuclear extracts and by the catalytic domain of the yeast enzyme Ulp1 (Figure 5.4C). In the case of the protease activity endogenous to HeLa nuclear extracts, cleavage of SUMO from recombinant MBD1 and the resistance to cleavage of the native protein can be observed in the same reaction. Therefore this experiment contains internal controls thus demonstrating an intrinsic difference between these two substrates rather than a spurious change in the reaction conditions used.

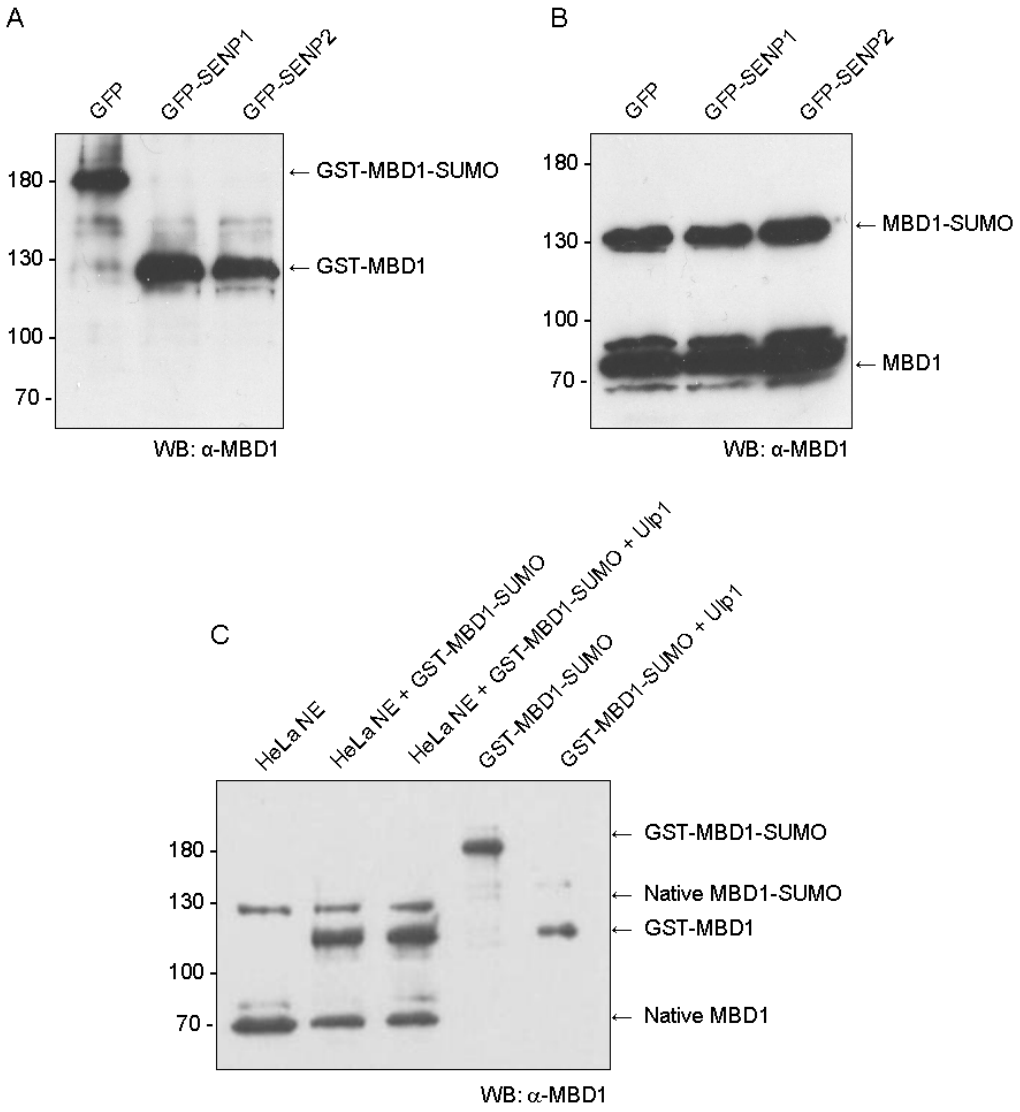

Figure 5.4 – Native SUMO modified MBD1 resists SUMO proteases *in vitro*

(A) GFP, GFP-SENP1 and GFP-SENP2 were expressed in HeLa cells and purified using an antibody against GFP. The precipitated proteins were incubated with recombinant SUMO modified GST-MBD1 and the reaction products were analysed by Western blot.
(B) The same purified proteins were incubated with HeLa nuclear extracts containing MBD1 and MBD1-SUMO. The protease activities were assayed by Western blot with antibodies against MBD1.
(C) The SUMO protease activities in HeLa nuclear extracts and Ulp1 both readily cleave recombinant MBD1-SUMO but not native MBD1-SUMO *in vitro*.

These results suggest that GST-tagged MBD1 is not a very good substrate to use in the study of SUMO modification of MBD1. Incorrect folding of bacterially expressed MBD1 might provide an explanation for these results. Alternatively, in contrast to previous reports, it appears that an as yet unidentified isoform other than PCM1 is the most abundant isoform of MBD1 in HeLa cells (see section 5.7). This too could be an important part of the explanation for the results described above.

5.5 SUMO modification destabilizes native MBD1 in vitro

It has previously been shown that the amount of SUMO modified MBD1 detectable in HeLa nuclear extracts increases substantially when the cysteine protease inhibitor NEM is included in the extraction buffers (Lyst *et al*, 2006). However, despite being catalytically active when expressed in HeLa cells, neither GFP-SENP1 nor GFP-SENP2 is able to cleave SUMO from native MBD1. This raises the possibility that SUMO conjugated MBD1 is a substrate for a known SUMO protease that has previously been reported either to be localized in the cytoplasm or to be specific for SUMO-2/3 conjugates, and which was therefore not tested here. Equally plausible is that MBD1 might be a substrate for an as yet uncharacterized protease. One method, which can be employed to discover such enzymes, is the chromatographic fractionation of extracts showing an *in vitro* activity followed by identification of the proteins involved by mass spectrometry. With this in mind, an assay was devised for the activity in HeLa nuclear extracts that degrades native SUMO modified MBD1.

In order to establish a useful assay, native SUMO modified MBD1 needed to be rapidly purified away from the unmodified protein as well as the endogenous protease activity of interest. Removal of the unmodified protein would allow the progression of the reaction to be monitored by the production of unconjugated MBD1 as well as the disappearance of the SUMO modified version. This separation was enabled by the differential extractability of unmodified and SUMO conjugated MBD1 from HeLa nuclei at physiological ionic strength (Figure 4.1A). After pre-extraction with buffer

containing 200 mM NaCl HeLa nuclear proteins were extracted from the pellet by sonication in the same buffer thus generating a protein fraction that was enriched in SUMO modified MBD1, but largely devoid of the unmodified protein. Separating modified MBD1 away from endogenous SUMO protease activities would improve any assay using this protein as a substrate because a greater proportion of any cleavage observed would be attributable to the fractions under scrutiny. This purification was achieved based on the affinity of MBD1 and SUMO modified MBD1 for immobilized nickel ions that had been uncovered during early attempts to purify a native MBD1-containing protein complex. Thus having partially purified native SUMO modified MBD1 by manipulation of the nuclear extraction protocol followed by immobilized metal affinity chromatography, a suitable substrate for the desired assay was in hand. Were a conventional Dignam HeLa nuclear extract to be used as a source of protease activity for this assay, then the analysis would be complicated by the presence of MBD1, both SUMO modified and unmodified, in this material. To circumvent this problem the flow-through after the application of a Dignam nuclear extract to a nickel column was selected as a starting material for a potential fractionation. This fraction was designated 'Ni^{2+}-FT' and contains most of the proteins found in the original extract, but is strongly depleted for both MBD1 and SUMO modified MBD1 (Figure 5.5).

To test the usefulness of these protein fractions in an assay for SUMO protease activity the 'Ni^{2+}-FT' fraction was incubated with the native SUMO modified MBD1 enriched extract and then the levels of the modified and unmodified protein were analysed by Western blotting with antibodies against MBD1 (Figure 5.5). The efficient separation of SUMO modified and unmodified MBD1 into nuclear pellet and extract fractions is shown (compare lanes 1 and 3) as is the ability of an Ni^{2+} ion containing column to deplete MBD1 from extracts (compare lanes 1 and 2). However, although treatment of native SUMO modified MBD1 with 'Ni^{2+}-FT' leads to the destruction of this protein, this is not accompanied by the appearance of unmodified MBD1 (lane 4). This effect is attributable to an activity in 'Ni^{2+}-FT' as the levels of SUMO modified MBD1 are not affected in this way when they are incubated with buffer alone in a control reaction in

parallel (lane 3). From these experiments it can be concluded that SUMO modification destabilizes MBD1 *in vitro* and that no activity exists in HeLa nuclear extracts capable of removing SUMO from MBD1 at an appreciable rate under the conditions tested.

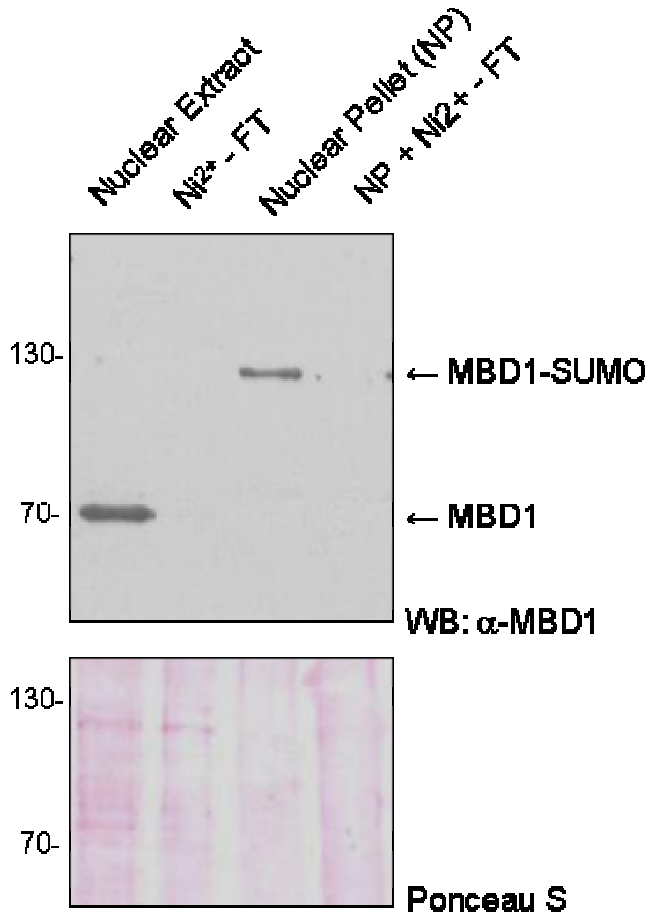

Figure 5.5 – Destabilization of native SUMO modified MBD1 *in vitro*

Western blot reveals that fractionation of HeLa nuclear proteins into nuclear extract and nuclear pellet efficiently separates SUMO modified and unmodified MBD1 (compare lanes 1 and 3). Nuclear extracts can be efficiently depleted of MBD1 taking the flow-through from a nickel column (lane 2). Native SUMO modified MBD1 is destroyed by an activity in HeLa nuclear extracts without being converted to detectable unmodified MBD1 (compare lane 3 with buffer alone to lane 4 where the 'Ni^{2+}-FT' fraction was added). Ponceau staining confirms approximately equal protein loading of the lanes.

5.6 MBD1-SUMO is not degraded by the proteasome in vivo

Recent reports have shown that modification of PML (promyelocytic leukaemia) protein with SUMO-2/3 is a signal for ubiquitination by the E3 ligase RNF4 and subsequent proteasomal degradation (Tatham *et al*, 2008; Lallemand-Breitenbach *et al*, 2008). These works also presented evidence that SUMO-1 can be a substrate for modification with SUMO-2/3 and that this might reflect a mechanism for the proteasomal targeting of some proteins modified with SUMO-1. Given that native SUMO modified MBD1 seems to be preferentially degraded in extracts without conversion to the unmodified form of the protein, it was hypothesised that SUMO modification of MBD1 might target the protein for proteasomal degradation, perhaps via ubiquitination by RNF4. It was therefore asked whether SUMO modified MBD1 might be specifically degraded by the proteasome *in vivo*.

The proteasome can be inhibited in cells using the drug MG132 (Lee and Goldberg, 1998). If SUMO modified MBD1 is targeted for ubiquitin-mediated degradation by the proteasome then following treatment of HeLa cells with this drug an accumulation of SUMO modified or other high molecular weight forms of MBD1 would be predicted. Cells were treated with MG132 or vehicle (DMSO) over a time course and whole cell extracts prepared under denaturing conditions were then analysed by SDS-PAGE followed by Western blotting with antibodies against either SUMO-2/3 or MBD1. Consistent with previous reports MG132 treatment, but not control treatment with DMSO, led to an increase in high molecular weight SUMO-2/3 conjugates (Figure 5.6A). This indicates successful inhibition of the proteasome in the experiment described here. However, no alteration was observed in the profile of MBD1 revealed by Western blotting (Figure 5.6B) suggesting that SUMO modification does not target MBD1 for degradation by the proteasome *in vivo*

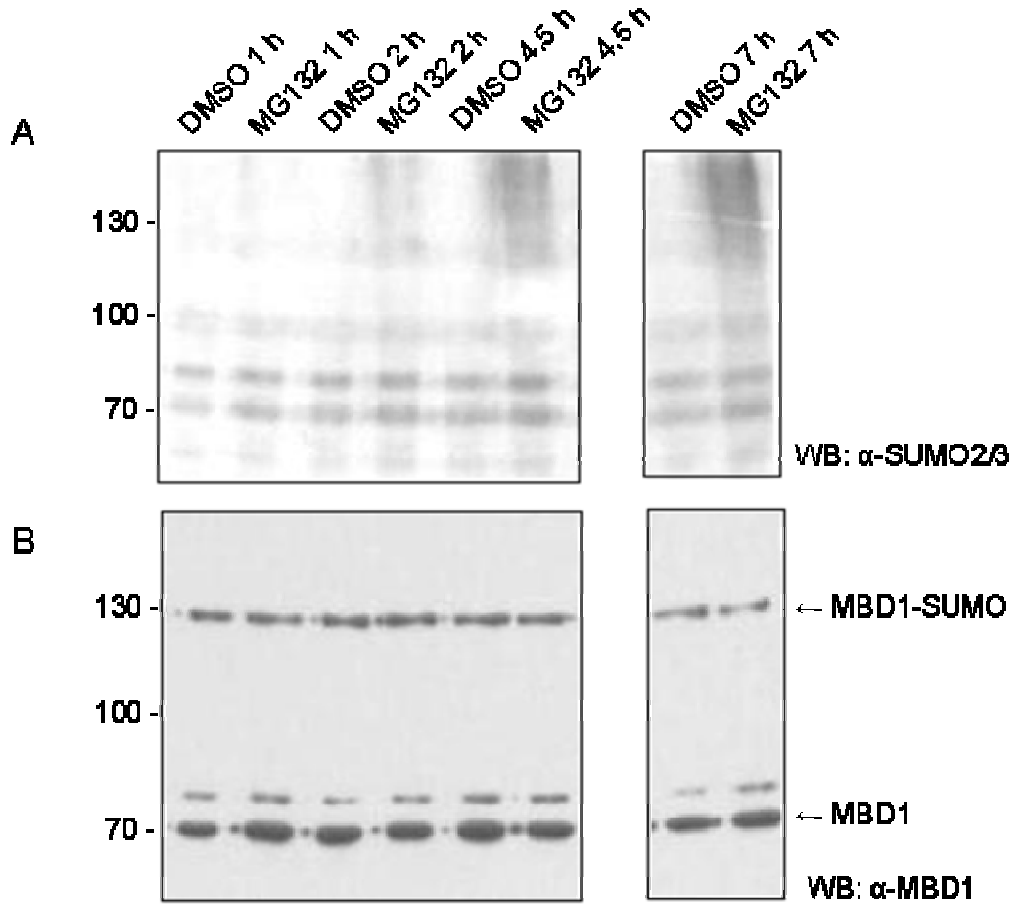

Figure 5.6 – SUMO modified MBD1 is not targeted by the proteasome *in vivo*

(A) HeLa cells were treated with either DMSO or MG132 for defined amounts of time. Western blot analysis whole cell extracts using antibodies against SUMO-2/3 showed increasing levels of high molecular weight SUMO-2/3 conjugates over time specifically when the cells were treated with MG132.
(B) The same membrane probed with an antibody against MBD1. Treatment of HeLa cells with MG132 does not lead to accumulation of MBD1 or SUMO modified MBD1.

5.7 PCM1 is not the major isoform of MBD1 in HeLa cells

There are various possible explanations for the different behaviour observed for recombinant SUMO modified MBD1 with respect to the native protein. It is possible that bacteria do not possess a chaperone that is required for proper folding of full length MBD1 or perhaps the purification scheme somehow disrupts the normal structure of the protein. An alternative explanation considered was that the isoform of MBD1 expressed in HeLa cells might not be the one, namely PCM1, which was expressed in and purified from *E.coli*. Part of the reason for entertaining this possibility was the fact that tagged forms of MBD1 used in this study consistently migrated slightly more slowly in SDS-PAGE than would be expected by simply adding the size of the tag to that of the endogenous MBD1 protein observed in HeLa cells. Furthermore, SUMO modification of the FS-MBD1 protein (PCM1 isoform described in chapters 3 and 4) was never observed in extracts even when modification of the endogenous protein was easily detectable.

Five isoforms of human MBD1 have been previously described in the literature (Cross *et al*, 1997; Fujita *et al*, 1999). These isoforms vary in sequence at the C-terminus as well as by which zinc finger domains are present in the central region of the protein (Figure 5.7A). Previously the major isoform of MBD1 in HeLa cells was reported to be PCM1 (Sarraf and Stancheva, 2004). This claim was based on RT-PCR analysis using primers flanking the first zinc finger of MBD1, in order that different sized PCR products would be observed depending on the presence or absence of this domain. In these experiments only a product corresponding to mRNA without this region was observed, and so it was concluded that PCM1, the only MBD1 isoform lacking this domain, must be the major variant of MBD1 present in HeLa cells. However, this interpretation assumed approximately equal amplification efficiencies of differently sized PCR products and therefore may not have been valid. This line of experimentation was therefore revisited and control PCR reactions using plasmid DNA as template were carried out in order to compare the amplification efficiencies of these PCR products. Both variant 3 and PCM1

isoforms of MBD1 amplified efficiently under the conditions previously used (Figure 5.7B) so it is unlikely that the conclusion that most MBD1 transcripts in HeLa code for a protein lacking the first zinc finger is erroneous.

Given the evidence that PCM1 is the most abundant isoform of MBD1 at the RNA level it might be expected that this is the most abundant version at the protein level too. However, recombinant untagged MBD1, and untagged MBD1 expressed in HeLa cells both migrate differently to the endogenous protein when assayed by SDS-PAGE followed by Western blotting (Figure 5.7C) with an antibody previously shown to be specific for MBD1 by siRNA mediated depletion of the protein (Sarraf and Stancheva, 2004). An earlier study using a different antibody found endogenous MBD1 migrated similarly to recombinant MBD1 in SDS-PAGE (Ng *et al*, 2000) but this result must now be treated with caution given that the specificity of the antibody was not verified by genetic means. Therefore the most abundant form of MBD1 in HeLa cells is likely a novel isoform that lacks the first zinc finger but is not PCM1. This has important implications for future research conducted on many aspects of the function of MBD1.

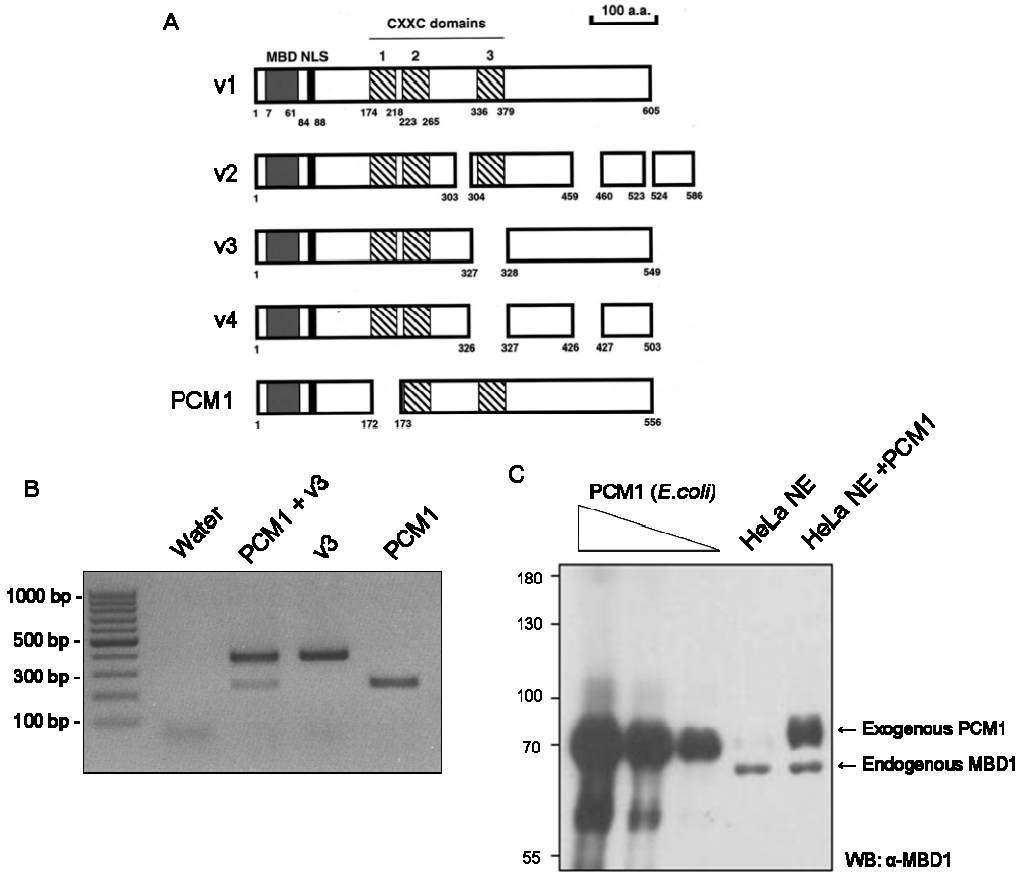

Figure 5.7 – An uncharacterized isoform accounts for most of the MBD1 protein in HeLa cells

(A) Schematic of the domains present in different MBD1 isoforms (adapted from Fujita *et al*, 1999). The only known isoform lacking the first zinc finger domain is PCM1.
(B) PCR using primers upstream and downstream of the first zinc finger of human MBD1 with the templates being PCM1-conaining and/or MBD1 variant 3 containing plasmid DNA. The longer product corresponds to variant 3 and the shorter product to PCM1. No bias is observed towards amplification of the shorter sequence.
(C) Western blot showing untagged bacterially expressed MBD1 (isoform PCM1) next to HeLa nuclear extracts containing only endogenous MBD1 and HeLa nuclear extracts over-expressing untagged MBD1 (isoform PCM1).

6. Chapter Six - Discussion

6.1 MBD1 potentially forms dimers

Up until now a detailed study of the biophysical properties of MBD1 has been lacking. Whereas MeCP2 behaves as an elongated monomer in nuclear extracts (Klose and Bird, 2004) and MBD2 is a component of the NuRD complex (Zhang *et al*, 1999; Feng and Zhang, 2001; Le Guezennec *et al*, 2006) this aspect of MBD1 function remained to be addressed. The starting point of the work in this thesis, therefore, was to ask whether or not MBD1 is a component of a stable multi-subunit protein complex. A number of lines of evidence presented in chapter three suggest that MBD1 does not stably interact with other proteins, but might self associate to form dimers. Firstly, the molecular mass of endogenous MBD1 derived from its Stoke's radius and sedimentation coefficient is close to double the theoretical molecular mass of this protein. Secondly, 3xFLAG-tagged MBD1 expressed in HeLa cells recapitulates these properties and after purification to near homogeneity the S value of this protein is observed to be unchanged. Thirdly, untagged full-length over-expressed MBD1 co-immunoprecipitates with an over-expressed 3xFLAG-tagged version of the protein (amino acids 1-395). Finally, recent experiments by a rotation student indicate that 3xFLAG-tagged MBD1 produced in rabbit reticulocyte lysates interacts with similarly expressed MYC-tagged MBD1. Using these *in vitro* translated proteins the region of MBD1 responsible for its putative self association was mapped to the zinc finger domains (Alex Tuck, unpublished data). This may then represent a similar situation to the one reported for the transcription factor Ikaros, which forms dimers via its zinc fingers (McCarty *et al*, 2003).

Despite these results doubt still remains over the conclusion that MBD1 self associates, and further work is required to test this hypothesis more rigorously. Although the FS-MBD1 protein used here is indistinguishable to endogenous MBD1 in terms of its Stoke's radius and sedimentation coefficient, a number of differences have still been

noted between these two forms of the protein, and so caution must be exercised when using FS-MBD1 as a model. For example, no interaction was found between FS-MBD1 and SETDB1, and SUMO modification of FS-MBD1 was not detected in extracts where modification of the endogenous protein was readily observed. Also, although FS-MBD1 migrates similarly to the endogenous protein in sucrose gradients at physiological ionic strength, the use of 420 mM sodium chloride leads to FS-MBD1 behaving like the bacterially expressed protein (like a monomer). This is not the case for endogenous MBD1, which maintains its S value even at the higher salt concentration (data not shown). Furthermore, given that a proportion of purified FS-MBD1 forms an aggregate as assessed by sucrose gradient sedimentation, it is not clear whether the co-immunoprecipitation data can be confidently interpreted as supporting the existence of MBD1 dimers. Another cause for concern is the fact that only over-expressed MBD1 co-immunoprecipitates with a 3xFLAG tagged version of the protein whilst the endogenous form remains unbound. A particularly useful experiment in resolving whether or not MBD1 forms dimers would be size exclusion analysis of a homogeneous preparation of FS-MBD1.

If the conclusion that MBD1 forms dimers is substantiated then it will be of interest to ask what the physiological significance of this self association might be. As the TRD appears to be dispensable for the interaction it seems unlikely that the formation of dimers is required for transcriptional repression *per se* (although it cannot be completely ruled out that self association was not 'rescued' by the Gal4 DNA binding domain used in reporter assays to map the TRD). An alternative possibility is that self association confers DNA sequence binding specificity and is thus required for correct targeting of MBD1 to specific parts of the genome. The appropriate spacing of two methyl binding domains and two non-methyl CpG binding zinc fingers could lead to the protein having high affinity for sites with a certain spacing of methylated and non-methylated CpGs. Another suggestion is that MBD1 self association might bring together different regions of genomic DNA and so lead to the formation of higher order chromatin structures and

compaction of the genome. A potentially fruitful line of future research might involve examining what effect SUMO modification has on self association of MBD1.

6.2 Binding partners of MBD proteins

Chapter four described experiments designed to identify interaction partners for MBD1, MeCP2, MBD2 and LSH that might have been missed by the approaches taken in previous studies. The method employed here utilized benzonase nuclease to solubilize these chromatin associated proteins under less stringent conditions than had been used in earlier works (Klose and Bird, 2004; Myant and Stancheva, 2008; Le Guezennec *et al*, 2006). Over-expressed 3xFLAG-tagged fusion proteins were purified using an antibody against this epitope, and co-purifying proteins were identified by mass spectrometry.

On the one hand the screen seems to have been successful in that a number of known interaction partners were re-discovered for all three MBD-containing proteins examined, and at the same time several putative novel binding proteins were identified. Also the stringent filtering criteria, which excluded any factors co-purifying with more than one of the bait proteins make the occurrence of false positives due to non-specific interactions with the tag or purification matrix used unlikely. Furthermore, Western blot analysis of samples prepared similarly revealed the data to be relatively robust and reproducible by an independent technique. On the other hand, however, the histone H3 lysine K9 methyltransferase SETDB1 as well as the chromatin assembly factor CAF1 were notable by their absence in the list of proteins co-purifying with MBD1. Given the weight of evidence for binding between MBD1 and SETDB1 (yeast two-hybrid, binding of recombinant proteins *in vitro* and reciprocal stoichiometric co-immunoprecipitation of endogenous proteins), the results of any experiment which is apparently inconsistent with this interaction must be treated with caution. It should also be kept in mind that the experiment described here is potentially prone to false negatives. For instance, it is certainly possible that not all of the interaction partners of these proteins are expressed in HeLa cells. It also cannot be ruled out that some proteins genuinely do interact with

more than one of the baits investigated here, and that such proteins would have been ignored due to attempts to guard against false positives. Indeed, this has clearly occurred with the histone deacetylases HDAC1 and HDAC2, which are known to interact with both MeCP2 and MBD2.

Co-immunoprecipitation between endogenous proteins is often viewed as a 'gold standard' for the verification of protein-protein interactions (Mackay *et al*, 2007). This type of experiment would be required to confirm that the putative interactions uncovered here do not represent over-expression artefacts. It will also be necessary to test the possibility that these candidate interactions are simply mediated by the fragments of DNA generated during generation of the extracts. Although the identification of proteins as co-purifying uniquely with one of the bait proteins goes some way to excluding this possibility, it is still plausible that the proteins examined here are bound to different genomic regions whose chromatin is composed of different complements of proteins. Testing for binding *in vitro* between recombinant proteins would go some way to ruling out the possibility of nucleosomal DNA mediated interactions and an additional approach currently being pursued is to repeat the screen using proteins harbouring point mutations such that they can no longer interact with DNA.

6.3 Regulation of SUMO modification of MBD1

Chapter five of this thesis focussed on SUMO modification of MBD1 and in particular whether this might be regulated by SUMO protease activities. Native SUMO modified MBD1 was found to be refractory to all such activities tested both *in vitro* and *in vivo*. The sensitivity of this protein to the absence of NEM in extraction buffers was confirmed, but surprisingly this appeared to be due to degradation of the protein rather than deconjugation of SUMO. The fact that the presence of SUMO conjugated proteins is often dependent on the presence of NEM or denaturing agents in extraction buffers is generally presumed to be due to the activity of SUMO proteases, but the data on MBD1 suggest that for some SUMO substrates at least, this might not be the case. The results

presented here are perhaps slightly surprising given that SUMO was readily cleaved form recombinant MBD1 by a variety of protease activities. One viewpoint is that the recombinant protein is misfolded and that this explains its different behaviour to native MBD1. However, another potentially more interesting resolution of the data is that different isoforms of MBD1 could show differential susceptibility to SUMO protease activities. Consistent with this notion is the fact that SUMO modification of over-expressed FS-MBD1 was not observed even in extracts where native SUMO modified MBD1 conjugates were present. This situation could be explained if SUMO is more readily cleaved from the PCM1 isoform than whichever version of MBD1 represents most of the protein in HeLa cells.

An urgent priority in future research on MBD1 will be to determine which isoforms are most abundantly expressed at the protein level in the tissues and cell lines under examination. Evidence is presented here that PCM1 behaves differently to the most abundant isoform of MBD1 present in HeLa cells in a variety of assays. This is not limited to behaviour with respect to SUMO modification, but also includes interaction with SETDB1 and susceptibility of putative self association to increasing ionic strength. If over-expression of MBD1 accounted for these changes then an alteration in the behaviour of the endogenous protein in the same extracts would be predicted. This, however, is not observed, and so a strong case can be argued for different isoforms of MBD1 being biochemically distinct. If it proves impossible to identify the most abundantly expressed MBD1 isoform, then it would be prudent to abandon approaches to the study of this protein which involve its over-expression.

6.4 Conclusions and future perspectives

Twenty years on since the identification of the first methyl-CpG binding activity (Meehan *et al*, 1989) the field has moved some way towards understanding how the DNA methylation signal is interpreted by the cell. Genetic studies in mice have revealed phenotypic abnormalities that occur when the proteins which specifically recognise

methylated DNA are ablated. In the case of Mbd2, there has been a degree of success in uncovering the molecular details of how this protein's absence leads to defects in the immune system and large intestine (Hutchins *et al*, 2002; Berger *et al*, 2007). Without Mbd2, specific genes seem to be mis-expressed concomitant with hyperacetylation of histones at their promoters. This is consistent with earlier work which identified MBD2 as a component of the NuRD complex (Zhang *et al*, 1999) and showed it to function as a histone deacetylase-dependent transcriptional repressor in reporter assays (Ng *et al*, 1999). With Mbd1 and Mecp2, however, it is still almost completely unknown how their absence leads to the neurological phenotypes observed in the null mice (Zhao *et al*, 2003; Chen *et al*, 2001; Guy *et al*, 2001). Although early work using reporter assays suggested that these proteins function as transcriptional repressors (Fujita *et al*, 1999; Nan *et al*, 1997), this has not been substantiated by expression analysis of *Mbd1*- or *Mecp2*-null mice (Zhao *et al*, 2003; Tudor *et al*, 2002). Thus little is fully understood about the physiological roles of these proteins.

The identification of binding partners can be useful in forming hypotheses as to what the functions of poorly characterized proteins might be. The tentative finding that MBD1 forms dimers is of some interest in this respect. For example, it is possible to imagine that MBD1 self association leads to the appropriate positioning of DNA binding domains such that a consensus binding sequence which can direct the protein to specific regions of the genome exists. In the light of this, chromatin immunoprecipitation experiments to identify loci occupied by MBD1 are a particularly pertinent avenue of research. In the case of MeCP2, the identification of a putative interaction with the PAF complex (Chaudhary *et al*, 2007) is arguably the most interesting finding here. If this interaction is borne out then it suggests a potential role for MeCP2 in the regulation of transcriptional elongation and mRNA processing that would merit further investigation. Perhaps then, the key to understanding the molecular defects in Rett syndrome and *Mecp2*-null mice is to let go of the idea that MeCP2 functions primarily as a transcriptional repressor of specific genes.

Appendix I – Plasmids used in this study

pET42a-MBD1 – Used to express MBD1 free of tags in *E.coli*. Full length MBD1 (isoform PCM1) was cloned with a C-terminal stop codon between the NdeI and XhoI sites of pET42a.

pBABE HYGRO MBD1-SSH-GFP – Used to make retroviruses for expression of MBD1 with C-terminal SBP-tag, S-tag, histidine hexamer and GFP-tag. The coding sequences were joined by PCR mediated recombination before being cloned into the EcoRI and BamHI sites of pBABE HYGRO.

pBABE HYGRO SSH-GFP – Used to make retrovirus for expression of an SBP-tag, S-tag and histidine hexamer with a C-terminal GFP. Generated using the same strategy as for pBABE HYGRO MBD1-SSH-GFP except the sequence of MBD1 was not included.

pF-MBD1-SSH – Used to express MBD1 with an N-terminal 3xFLAG tag and C-terminal SBP-tag, S-tag and histidine hexamer. The SSH tag was cloned into the XbaI site of p3xFLAG-CMV-10. MBD1 was then cloned between the NotI and BglII sites.

pFS-MBD1 – Used to express MBD1 with N-terminal 3xFLAG- and S-tags in HeLa cells. The backbone vector was produced by cloning an S-tag between the HindIII and NotI sites of p3xFLAG-CMV-10. Full length MBD1 (isoform PCM1) was cloned between the NotI and BglII sites of the resulting plasmid.

pFS-LSH – Same as pFS-MBD1 except LSH coding sequence was used instead.

pFS-MeCP2 – Same as pFS-MBD1 except MeCP2 coding sequence was used instead.

pF-Mbd2 – A gift from Aimee Deaton. The mouse *mbd2* sequence is cloned into p3xFLAG-CMV-10 between the NotI and BglII sites.

pFS-MBD1(1-395) – Same as pFS-MBD1 except only the region coding for amino acids 1-395 was cloned.

pGST-MBD1 - Used to express MBD1 as a GST fusion in *E.coli*. Full length MBD1 (isoform PCM1) was cloned between the EcoRI and SalI sites of pGEX4t1.

pBADE12 – A gift from Mario Mencía (see Mencía and de Lorenzo, 2004). Used to express the SUMO conjugation enzymes in *E.coli*.

pRSFDuet1-SUMO-1 – Used to express SUMO-1 in *E.coli* with an N-terminal histidine hexamer. Full-length SUMO-1 up to the C-terminal diglycine motif was cloned between the BamHI and HindIII sites of pRSFDuet1.

pEGFP-C3-SENP1 – Used to express SENP1 as a GFP fusion in HeLa cells. The full length coding region was cloned between the HindIII and EcoRI sites of pEGFP-C3.

pEGFP-C3-SENP2 – Used to express SENP2 as a GFP fusion in HeLa cells. The full length coding region was cloned between the HindIII and EcoRI sites of pEGFP-C3.

pEGFP-C3 – The unmodified vector was used for the expression of GFP in HeLa cells.

pEGFP-N1-MBD1 – Used to express untagged MBD1 in HeLa cells. Full length MBD1 (isoform PCM1) was cloned with a C-terminal stop codon between the SacI and EcoRI sites of pEGFP-N1.

pFS-MBD1(Var3) – Same as pFS-MBD1 except a different isoform was cloned. Used as a PCR template to rule out differential amplification of PCM1 and variant 3 isoforms. The first zinc finger of MBD1 was amplified from HeLa cDNA and then used with PCM1 sequences in a PCR recombination mediated PCR strategy to generate variant 3.

Appendix II – proteins identified by mass spectrometry in FLAG immunoprecipitations

The proteins purified using the M2 anti-FLAG antibody and identified by mass spectrometry are listed in the following table. The samples proteins were found in are colour-coded as follows:

Non-transfected cells
FS-MBD1
FS-LSH
FS-MeCP2
F-mMBD2

Accession number	Peptides	Description
IPI00418234	43	Methyl-CpG-binding protein 2
IPI00003519	8	116 kDa U5 small nuclear ribonucleoprotein component
IPI00003519	4	116 kDa U5 small nuclear ribonucleoprotein component
IPI00816451	9	129 kDa protein
IPI00005625	11	136 kDa protein
IPI00456681	13	199 kDa protein
IPI00031627	8	217 kDa protein
IPI00031627	8	217 kDa protein
IPI00414482	7	241 kDa protein
IPI00018398	2	26S protease regulatory subunit 6A
IPI00021435	3	26S protease regulatory subunit 7
IPI00023919	2	26S protease regulatory subunit 8
IPI00021926	2	26S protease regulatory subunit S10B
IPI00021926	3	26S protease regulatory subunit S10B
IPI00012268	2	26S proteasome non-ATPase regulatory subunit 2
IPI00012268	6	26S proteasome non-ATPase regulatory subunit 2
IPI00011603	4	26S proteasome non-ATPase regulatory subunit 3
IPI00011603	6	26S proteasome non-ATPase regulatory subunit 3
IPI00793917	1	27 kDa protein
IPI00293845	13	275 kDa protein
IPI00293845	7	275 kDa protein
IPI00293845	13	275 kDa protein
IPI00294242	6	28S ribosomal protein S31, mitochondrial precursor
IPI00294242	8	28S ribosomal protein S31, mitochondrial precursor
IPI00073779	4	28S ribosomal protein S35, mitochondrial precursor
IPI00073779	10	28S ribosomal protein S35, mitochondrial precursor
IPI00791156	3	31 kDa protein
IPI00791156	1	31 kDa protein

IPI00555610	8	313 kDa protein
IPI00163866	5	330 kDa protein
IPI00023086	9	39S ribosomal protein L15, mitochondrial precursor
IPI00027096	9	39S ribosomal protein L19, mitochondrial precursor
IPI00411816	3	39S ribosomal protein L2, mitochondrial precursor
IPI00162330	2	39S ribosomal protein L37, mitochondrial precursor
IPI00162330	11	39S ribosomal protein L37, mitochondrial precursor
IPI00783656	9	39S ribosomal protein L38, mitochondrial precursor
IPI00009680	10	39S ribosomal protein L44, mitochondrial precursor
IPI00185859	5	39S ribosomal protein L45, mitochondrial precursor
IPI00023161	6	39S ribosomal protein L46, mitochondrial precursor
IPI00008438	5	40S ribosomal protein S10
IPI00216153	1	40S ribosomal protein S15
IPI00221093	3	40S ribosomal protein S17
IPI00013485	7	40S ribosomal protein S2
IPI00011253	5	40S ribosomal protein S3
IPI00011253	8	40S ribosomal protein S3
IPI00419880	6	40S ribosomal protein S3a
IPI00217030	13	40S ribosomal protein S4, X isoform
IPI00021840	8	40S ribosomal protein S6
IPI00013415	5	40S ribosomal protein S7
IPI00216587	2	40S ribosomal protein S8
IPI00216587	7	40S ribosomal protein S8
IPI00553164	8	40S ribosomal protein SA
IPI00027493	3	4F2 cell-surface antigen heavy chain
IPI00383111	19	57 kDa protein
IPI00383111	21	57 kDa protein
IPI00383111	14	57 kDa protein
IPI00383111	17	57 kDa protein
IPI00011253	15	40S ribosomal protein S3
IPI00419880	17	40S ribosomal protein S3a
IPI00383111	27	6 57 kDa protein
IPI00021439	15	Actin, cytoplasmic 1
IPI00059366	18	H2A histone family, member Y isoform 2
IPI00395865	16	Histone-binding protein RBBP7
IPI00215734	21	HRMT1L2 protein
IPI00438701	32	Isoform 1 of Methyl-CpG-binding domain protein 1
IPI00434623	22	Isoform 1 of Methyl-CpG-binding domain protein 2
IPI00549248	12	Isoform 1 of Nucleophosmin
IPI00216592	14	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2
IPI00009865	1	Keratin, type I cytoskeletal 10
IPI00019359	25	Keratin, type I cytoskeletal 9
IPI00220327	38	Keratin, type II cytoskeletal 1
IPI00220327	36	Keratin, type II cytoskeletal 1
IPI00021304	34	Keratin, type II cytoskeletal 2 epidermal
IPI00641950	18	Lung cancer oncogene 7
IPI00171798	34	Metastasis-associated protein MTA2
IPI00449049	42	Poly [ADP-ribose] polymerase 1

IPI00010590	88	606 Isoform 1 of Lymphoid-specific helicase
IPI00008530	5	60S acidic ribosomal protein P0
IPI00008530	4	60S acidic ribosomal protein P0
IPI00008530	6	60S acidic ribosomal protein P0
IPI00008530	11	60S acidic ribosomal protein P0
IPI00554723	3	60S ribosomal protein L10
IPI00412579	4	60S ribosomal protein L10a
IPI00215719	2	60S ribosomal protein L18
IPI00215719	6	60S ribosomal protein L18
IPI00025329	4	60S ribosomal protein L19
IPI00306332	4	60S ribosomal protein L24
IPI00550021	11	60S ribosomal protein L3
IPI00003918	9	60S ribosomal protein L4
IPI00003918	4	60S ribosomal protein L4
IPI00003918	8	60S ribosomal protein L4
IPI00003918	15	60S ribosomal protein L4
IPI00000494	4	60S ribosomal protein L5
IPI00000494	7	60S ribosomal protein L5
IPI00329389	1	60S ribosomal protein L6
IPI00329389	4	60S ribosomal protein L6
IPI00329389	7	60S ribosomal protein L6
IPI00030179	10	60S ribosomal protein L7
IPI00299573	13	60S ribosomal protein L7a
IPI00031691	3	60S ribosomal protein L9
IPI00472102	6	61 kDa protein
IPI00472102	16	61 kDa protein
IPI00465430	9	70 kDa protein
IPI00465430	4	70 kDa protein
IPI00465430	12	70 kDa protein
IPI00478128	16	70 kDa protein
IPI00465430	5	70 kDa protein
IPI00181231	10	79 kDa protein
IPI00334775	3	85 kDa protein
IPI00797796	5	92 kDa protein
IPI00025849	3	Acidic leucine-rich nuclear phosphoprotein 32 family member A
IPI00021439	10	Actin, cytoplasmic 1
IPI00021439	13	Actin, cytoplasmic 1
IPI00021439	7	Actin, cytoplasmic 1
IPI00021439	12	Actin, cytoplasmic 1
IPI00022215	11	Activity-dependent neuroprotector
IPI00022215	15	Activity-dependent neuroprotector
IPI00022215	7	Activity-dependent neuroprotector
IPI00033907	6	Anaphase-promoting complex subunit 1
IPI00008248	3	Anaphase-promoting complex subunit 7
IPI00008248	4	Anaphase-promoting complex subunit 7
IPI00159899	4	ankyrin repeat and FYVE domain containing 1 isoform 1
IPI00418169	5	annexin A2 isoform 1
IPI00470657	5	Anti-colorectal carcinoma heavy chain

IPI00470657	8	Anti-colorectal carcinoma heavy chain
IPI00470657	7	Anti-colorectal carcinoma heavy chain
IPI00470657	6	Anti-colorectal carcinoma heavy chain
IPI00470657	5	Anti-colorectal carcinoma heavy chain
IPI00556297	2	Arginine/serine-rich splicing factor 6 variant (Fragment)
IPI00328828	7	AT rich interactive domain 4B isoform 1
IPI00220834	8	ATP-dependent DNA helicase 2 subunit 2
IPI00220834	15	ATP-dependent DNA helicase 2 subunit 2
IPI00742905	9	ATP-dependent RNA helicase A
IPI00742905	14	ATP-dependent RNA helicase A
IPI00742905	11	ATP-dependent RNA helicase A
IPI00742905	17	ATP-dependent RNA helicase A
IPI00742905	25	ATP-dependent RNA helicase A
IPI00006987	7	ATP-dependent RNA helicase DDX24
IPI00031508	2	ATP-dependent RNA helicase DHX8
IPI00290684	6	Bifunctional polynucleotide phosphatase/kinase
IPI00254408	6	bromodomain PHD finger transcription factor isoform 1
IPI00328987	9	Bystin
IPI00333010	3	calcium homeostasis endoplasmic reticulum protein
IPI00333010	5	calcium homeostasis endoplasmic reticulum protein
IPI00220959	3	Cartilage-associated protein precursor
IPI00020602	7	Casein kinase II subunit alpha'
IPI00444262	5	CDNA FLJ45706 fis, clone FEBRA2028457, highly similar to Nucleolin
IPI00444262	16	CDNA FLJ45706 fis, clone FEBRA2028457, highly similar to Nucleolin
IPI00444262	8	CDNA FLJ45706 fis, clone FEBRA2028457, highly similar to Nucleolin
IPI00444262	26	CDNA FLJ45706 fis, clone FEBRA2028457, highly similar to Nucleolin
IPI00444262	17	CDNA FLJ45706 fis, clone FEBRA2028457, highly similar to Nucleolin
IPI00444046	3	CDNA FLJ46149 fis, clone TESTI4000621, moderately similar to Snf2-related CBP activator protein
IPI00465294	1	Cell division cycle 5-like protein
IPI00217357	4	Cell division cycle and apoptosis regulator protein 1
IPI00217357	7	Cell division cycle and apoptosis regulator protein 1
IPI00014264	3	Centaurin-beta 2
IPI00014264	4	Centaurin-beta 2
IPI00297579	4	Chromobox protein homolog 3
IPI00307659	3	Chromobox protein homolog 8
IPI00152535	1	Chromodomain helicase-DNA-binding protein 5
IPI00386043	4	Chromosome 1 open reading frame 27
IPI00386043	7	Chromosome 1 open reading frame 27
IPI00014230	7	Complement component 1 Q subcomponent-binding protein, mitochondrial precursor
IPI00008453	3	Coronin-1C
IPI00016613	9	CSNK2A1 protein
IPI00016613	4	CSNK2A1 protein
IPI00028275	1	Cytoskeleton-associated protein 5
IPI00396435	12	DEAH (Asp-Glu-Ala-His) box polypeptide 15
IPI00396435	7	DEAH (Asp-Glu-Ala-His) box polypeptide 15
IPI00396435	12	DEAH (Asp-Glu-Ala-His) box polypeptide 15
IPI00396435	5	DEAH (Asp-Glu-Ala-His) box polypeptide 15
IPI00025753	9	Desmoglein-1 precursor

IPI00025753	3	Desmoglein-1 precursor
IPI00465233	7	DJ1014D13.1 protein
IPI00465233	14	DJ1014D13.1 protein
IPI00465233	9	DJ1014D13.1 protein
IPI00465233	11	DJ1014D13.1 protein
IPI00465233	12	DJ1014D13.1 protein
IPI00329216	2	DNA (cytosine-5)-methyltransferase 3A
IPI00329216	6	DNA (cytosine-5)-methyltransferase 3A
IPI00293464	21	DNA damage-binding protein 1
IPI00293464	4	DNA damage-binding protein 1
IPI00293464	13	DNA damage-binding protein 1
IPI00293464	9	DNA damage-binding protein 1
IPI00184330	8	DNA replication licensing factor MCM2
IPI00184330	6	DNA replication licensing factor MCM2
IPI00013214	6	DNA replication licensing factor MCM3
IPI00013214	8	DNA replication licensing factor MCM3
IPI00018349	5	DNA replication licensing factor MCM4
IPI00018350	4	DNA replication licensing factor MCM5
IPI00413611	5	DNA topoisomerase 1
IPI00413611	11	DNA topoisomerase 1
IPI00027808	10	DNA-directed RNA polymerase II 140 kDa polypeptide
IPI00012535	1	DnaJ homolog subfamily A member 1
IPI00012535	7	DnaJ homolog subfamily A member 1
IPI00025874	3	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 67 kDa subunit precursor
IPI00003406	1	Drebrin
IPI00026993	3	E3 ubiquitin-protein ligase RING2
IPI00797945	15	E3 ubiquitin-protein ligase UHRF1
IPI00797945	16	E3 ubiquitin-protein ligase UHRF1
IPI00797945	8	E3 ubiquitin-protein ligase UHRF1
IPI00745955	7	EBNA1 binding protein 2
IPI00384472	1	EHMT1 protein
IPI00384472	1	EHMT1 protein
IPI00396485	7	Elongation factor 1-alpha 1
IPI00396485	8	Elongation factor 1-alpha 1
IPI00396485	9	Elongation factor 1-alpha 1
IPI00396485	5	Elongation factor 1-alpha 1
IPI00014424	2	Elongation factor 1-alpha 2
IPI00014424	1	Elongation factor 1-alpha 2
IPI00014424	2	Elongation factor 1-alpha 2
IPI00023048	4	Elongation factor 1-delta
IPI00000875	3	Elongation factor 1-gamma
IPI00000875	4	Elongation factor 1-gamma
IPI00000875	7	Elongation factor 1-gamma
IPI00186290	7	Elongation factor 2
IPI00186290	5	Elongation factor 2
IPI00186290	9	Elongation factor 2
IPI00186290	18	Elongation factor 2
IPI00171248	6	embryonic ectoderm development isoform b

IPI00376787	6	enhancer of zeste 2 isoform a
IPI00376787	6	enhancer of zeste 2 isoform a
IPI00025491	4	Eukaryotic initiation factor 4A-I
IPI00025491	12	Eukaryotic initiation factor 4A-I
IPI00297982	6	Eukaryotic translation initiation factor 2 subunit 3
IPI00290461	8	Eukaryotic translation initiation factor 3 subunit 1
IPI00029012	7	Eukaryotic translation initiation factor 3 subunit 10
IPI00029012	14	Eukaryotic translation initiation factor 3 subunit 10
IPI00029012	7	Eukaryotic translation initiation factor 3 subunit 10
IPI00029012	13	Eukaryotic translation initiation factor 3 subunit 10
IPI00029012	29	Eukaryotic translation initiation factor 3 subunit 10
IPI00012795	7	Eukaryotic translation initiation factor 3 subunit 2
IPI00012795	9	Eukaryotic translation initiation factor 3 subunit 2
IPI00647650	6	Eukaryotic translation initiation factor 3 subunit 3
IPI00647650	5	Eukaryotic translation initiation factor 3 subunit 3
IPI00290460	6	Eukaryotic translation initiation factor 3 subunit 4
IPI00654777	1	Eukaryotic translation initiation factor 3 subunit 5
IPI00654777	4	Eukaryotic translation initiation factor 3 subunit 5
IPI00654777	4	Eukaryotic translation initiation factor 3 subunit 5
IPI00654777	4	Eukaryotic translation initiation factor 3 subunit 5
IPI00654777	7	Eukaryotic translation initiation factor 3 subunit 5
IPI00013068	3	Eukaryotic translation initiation factor 3 subunit 6
IPI00013068	7	Eukaryotic translation initiation factor 3 subunit 6
IPI00013068	6	Eukaryotic translation initiation factor 3 subunit 6
IPI00013068	11	Eukaryotic translation initiation factor 3 subunit 6
IPI00013068	13	Eukaryotic translation initiation factor 3 subunit 6
IPI00006181	12	Eukaryotic translation initiation factor 3 subunit 7
IPI00006181	6	Eukaryotic translation initiation factor 3 subunit 7
IPI00006181	9	Eukaryotic translation initiation factor 3 subunit 7
IPI00006181	20	Eukaryotic translation initiation factor 3 subunit 7
IPI00016910	4	Eukaryotic translation initiation factor 3 subunit 8
IPI00016910	12	Eukaryotic translation initiation factor 3 subunit 8
IPI00016910	4	Eukaryotic translation initiation factor 3 subunit 8
IPI00016910	9	Eukaryotic translation initiation factor 3 subunit 8
IPI00016910	17	Eukaryotic translation initiation factor 3 subunit 8
IPI00012079	4	Eukaryotic translation initiation factor 4B
IPI00012079	3	Eukaryotic translation initiation factor 4B
IPI00439415	6	eukaryotic translation initiation factor 4B
IPI00012079	6	Eukaryotic translation initiation factor 4B
IPI00010105	5	Eukaryotic translation initiation factor 6
IPI00026970	16	FACT complex subunit SPT16
IPI00026970	6	FACT complex subunit SPT16
IPI00026970	18	FACT complex subunit SPT16
IPI00026970	8	FACT complex subunit SPT16
IPI00005154	8	FACT complex subunit SSRP1
IPI00005154	4	FACT complex subunit SSRP1
IPI00005154	12	FACT complex subunit SSRP1
IPI00005154	11	FACT complex subunit SSRP1

IPI00218132	4	F-box-like/WD repeat protein TBL1X
IPI00002922	13	F-box-like/WD repeat protein TBL1XR1
IPI00302592	89	filamin A, alpha
IPI00302592	80	filamin A, alpha
IPI00302592	88	filamin A, alpha
IPI00302592	58	filamin A, alpha
IPI00302592	72	filamin A, alpha
IPI00642018	8	FLYWCH-type zinc finger 1 isoform b
IPI00069750	3	fuse-binding protein-interacting repressor isoform a
IPI00102069	5	GA17 protein
IPI00102069	3	GA17 protein
IPI00016801	4	Glutamate dehydrogenase 1, mitochondrial precursor
IPI00219018	3	Glyceraldehyde-3-phosphate dehydrogenase
IPI00005132	6	guanine nucleotide binding protein-like 3 (nucleolar)-like
IPI00059366	13	H2A histone family, member Y isoform 2
IPI00059366	6	H2A histone family, member Y isoform 2
IPI00059366	14	H2A histone family, member Y isoform 2
IPI00304925	1	Heat shock 70 kDa protein 1
IPI00304925	10	Heat shock 70 kDa protein 1
IPI00304925	4	Heat shock 70 kDa protein 1
IPI00304925	2	Heat shock 70 kDa protein 1
IPI00304925	4	Heat shock 70 kDa protein 1
IPI00301277	1	Heat shock 70 kDa protein 1L
IPI00339269	1	Heat shock 70 kDa protein 6
IPI00788958	4	Heat shock 70kDa protein 9B variant (Fragment)
IPI00788958	6	Heat shock 70kDa protein 9B variant (Fragment)
IPI00465365	9	heterogeneous nuclear ribonucleoprotein A1 isoform a
IPI00215965	5	heterogeneous nuclear ribonucleoprotein A1 isoform b
IPI00215965	10	heterogeneous nuclear ribonucleoprotein A1 isoform b
IPI00011274	1	heterogeneous nuclear ribonucleoprotein D-like
IPI00003881	1	heterogeneous nuclear ribonucleoprotein F
IPI00003881	1	heterogeneous nuclear ribonucleoprotein F
IPI00304692	3	Heterogeneous nuclear ribonucleoprotein G
IPI00304692	3	Heterogeneous nuclear ribonucleoprotein G
IPI00304692	5	Heterogeneous nuclear ribonucleoprotein G
IPI00026230	1	Heterogeneous nuclear ribonucleoprotein H'
IPI00013881	4	heterogeneous nuclear ribonucleoprotein H1
IPI00013881	7	heterogeneous nuclear ribonucleoprotein H1
IPI00013881	5	heterogeneous nuclear ribonucleoprotein H1
IPI00013881	7	heterogeneous nuclear ribonucleoprotein H1
IPI00027834	4	heterogeneous nuclear ribonucleoprotein L isoform a
IPI00027834	8	heterogeneous nuclear ribonucleoprotein L isoform a
IPI00027834	5	heterogeneous nuclear ribonucleoprotein L isoform a
IPI00027834	8	heterogeneous nuclear ribonucleoprotein L isoform a
IPI00027834	10	heterogeneous nuclear ribonucleoprotein L isoform a
IPI00171903	4	heterogeneous nuclear ribonucleoprotein M isoform a
IPI00171903	9	heterogeneous nuclear ribonucleoprotein M isoform a
IPI00171903	7	heterogeneous nuclear ribonucleoprotein M isoform a

IPI00171903	7	heterogeneous nuclear ribonucleoprotein M isoform a
IPI00171903	8	heterogeneous nuclear ribonucleoprotein M isoform a
IPI00012074	5	Heterogeneous nuclear ribonucleoprotein R
IPI00012074	4	Heterogeneous nuclear ribonucleoprotein R
IPI00012074	9	Heterogeneous nuclear ribonucleoprotein R
IPI00012074	11	Heterogeneous nuclear ribonucleoprotein R
IPI00479217	6	heterogeneous nuclear ribonucleoprotein U isoform b
IPI00479217	7	heterogeneous nuclear ribonucleoprotein U isoform b
IPI00479217	10	heterogeneous nuclear ribonucleoprotein U isoform b
IPI00479217	12	heterogeneous nuclear ribonucleoprotein U isoform b
IPI00479217	10	heterogeneous nuclear ribonucleoprotein U isoform b
IPI00180764	4	Histone acetyltransferase MYST2
IPI00013774	4	Histone deacetylase 1
IPI00013774	12	Histone deacetylase 1
IPI00289601	3	histone deacetylase 2
IPI00289601	5	histone deacetylase 2
IPI00289601	9	histone deacetylase 2
IPI00289601	19	histone deacetylase 2
IPI00026272	1	Histone H2A type 1-B
IPI00026272	1	Histone H2A type 1-B
IPI00081836	1	Histone H2A type 1-H
IPI00081836	3	Histone H2A type 1-H
IPI00081836	3	Histone H2A type 1-H
IPI00081836	5	Histone H2A type 1-H
IPI00216457	5	Histone H2A type 2-A
IPI00003935	2	Histone H2B type 2-E
IPI00003935	1	Histone H2B type 2-E
IPI00794461	3	Histone H2B.d
IPI00465070	1	Histone H3.1
IPI00219038	4	Histone H3.3
IPI00453473	6	Histone H4
IPI00453473	4	Histone H4
IPI00453473	4	Histone H4
IPI00328319	2	Histone-binding protein RBBP4
IPI00328319	3	Histone-binding protein RBBP4
IPI00328319	11	Histone-binding protein RBBP4
IPI00395865	10	Histone-binding protein RBBP7
IPI00395865	2	Histone-binding protein RBBP7
IPI00395865	11	Histone-binding protein RBBP7
IPI00013788	3	HIV Tat-specific factor 1
IPI00013788	12	HIV Tat-specific factor 1
IPI00003362	7	HSPA5 protein
IPI00003362	14	HSPA5 protein
IPI00003362	8	HSPA5 protein
IPI00003362	15	HSPA5 protein
IPI00003362	7	HSPA5 protein
IPI00375731	18	Hypothetical protein DKFZp686E2459
IPI00375731	15	Hypothetical protein DKFZp686E2459

IPI00375731	20	Hypothetical protein DKFZp686E2459
IPI00375731	12	Hypothetical protein DKFZp686E2459
IPI00375731	12	Hypothetical protein DKFZp686E2459
IPI00026533	2	Hypothetical protein DKFZp686M21107
IPI00106955	7	Hypothetical protein LOC144097
IPI00106955	6	Hypothetical protein LOC144097
IPI00795088	7	Hypothetical protein XRCC5
IPI00397801	1	Ifapsoriasin
IPI00303292	1	Importin alpha-1 subunit
IPI00303292	3	Importin alpha-1 subunit
IPI00002214	1	Importin alpha-2 subunit
IPI00002214	6	Importin alpha-2 subunit
IPI00002214	7	Importin alpha-2 subunit
IPI00002214	12	Importin alpha-2 subunit
IPI00002214	7	Importin alpha-2 subunit
IPI00299033	6	Importin alpha-3 subunit
IPI00012578	2	Importin alpha-4 subunit
IPI00012578	10	Importin alpha-4 subunit
IPI00747764	3	Importin alpha-7 subunit
IPI00747764	2	Importin alpha-7 subunit
IPI00014319	3	influenza virus NS1A binding protein isoform a
IPI00291510	13	Inosine-5'-monophosphate dehydrogenase 2
IPI00291510	12	Inosine-5'-monophosphate dehydrogenase 2
IPI00005198	2	Interleukin enhancer-binding factor 2
IPI00005198	4	Interleukin enhancer-binding factor 2
IPI00005198	2	Interleukin enhancer-binding factor 2
IPI00005198	9	Interleukin enhancer-binding factor 2
IPI00299608	5	Isoform 1 of 26S proteasome non-ATPase regulatory subunit 1
IPI00004511	3	Isoform 1 of 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
IPI00004511	4	Isoform 1 of 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
IPI00007423	5	Isoform 1 of Acidic leucine-rich nuclear phosphoprotein 32 family member B
IPI00003627	2	Isoform 1 of Actin-like protein 6A
IPI00003627	4	Isoform 1 of Actin-like protein 6A
IPI00003627	2	Isoform 1 of Actin-like protein 6A
IPI00003627	3	Isoform 1 of Actin-like protein 6A
IPI00549736	6	Isoform 1 of Activating signal cointegrator 1 complex subunit 2
IPI00430472	8	Isoform 1 of Activating signal cointegrator 1 complex subunit 3
IPI00430472	12	Isoform 1 of Activating signal cointegrator 1 complex subunit 3
IPI00301503	3	Isoform 1 of Arginine/serine-rich-splicing factor 10
IPI00301503	2	Isoform 1 of Arginine/serine-rich-splicing factor 10
IPI00301503	2	Isoform 1 of Arginine/serine-rich-splicing factor 10
IPI00301503	4	Isoform 1 of Arginine/serine-rich-splicing factor 10
IPI00513797	3	Isoform 1 of Beta-catenin-like protein 1
IPI00296388	3	Isoform 1 of Bromodomain adjacent to zinc finger domain 2A
IPI00296388	10	Isoform 1 of Bromodomain adjacent to zinc finger domain 2A
IPI00412415	9	Isoform 1 of Bromodomain adjacent to zinc finger domain protein 1A
IPI00069817	5	Isoform 1 of Bromodomain adjacent to zinc finger domain protein 1B
IPI00069817	5	Isoform 1 of Bromodomain adjacent to zinc finger domain protein 1B

IPI00373870	1	Isoform 1 of Chromodomain helicase-DNA-binding protein 3
IPI00373870	3	Isoform 1 of Chromodomain helicase-DNA-binding protein 3
IPI00373870	9	Isoform 1 of Chromodomain helicase-DNA-binding protein 3
IPI00000846	29	Isoform 1 of Chromodomain helicase-DNA-binding protein 4
IPI00000846	3	Isoform 1 of Chromodomain helicase-DNA-binding protein 4
IPI00000846	47	Isoform 1 of Chromodomain helicase-DNA-binding protein 4
IPI00000846	64	Isoform 1 of Chromodomain helicase-DNA-binding protein 4
IPI00012593	5	Isoform 1 of DNA (cytosine-5)-methyltransferase 3B
IPI00021518	3	Isoform 1 of DNA damage-binding protein 2
IPI00021518	5	Isoform 1 of DNA damage-binding protein 2
IPI00031801	1	Isoform 1 of DNA-binding protein A
IPI00296337	8	Isoform 1 of DNA-dependent protein kinase catalytic subunit
IPI00296337	46	Isoform 1 of DNA-dependent protein kinase catalytic subunit
IPI00296337	16	Isoform 1 of DNA-dependent protein kinase catalytic subunit
IPI00296337	37	Isoform 1 of DNA-dependent protein kinase catalytic subunit
IPI00296337	11	Isoform 1 of DNA-dependent protein kinase catalytic subunit
IPI00394665	5	Isoform 1 of Double-stranded RNA-specific adenosine deaminase
IPI00394665	8	Isoform 1 of Double-stranded RNA-specific adenosine deaminase
IPI00396370	7	Isoform 1 of Eukaryotic translation initiation factor 3 subunit 9
IPI00396370	7	Isoform 1 of Eukaryotic translation initiation factor 3 subunit 9
IPI00396370	8	Isoform 1 of Eukaryotic translation initiation factor 3 subunit 9
IPI00396370	23	Isoform 1 of Eukaryotic translation initiation factor 3 subunit 9
IPI00289334	2	Isoform 1 of Filamin-B
IPI00289334	6	Isoform 1 of Filamin-B
IPI00003865	15	Isoform 1 of Heat shock cognate 71 kDa protein
IPI00003865	17	Isoform 1 of Heat shock cognate 71 kDa protein
IPI00003865	16	Isoform 1 of Heat shock cognate 71 kDa protein
IPI00003865	17	Isoform 1 of Heat shock cognate 71 kDa protein
IPI00003865	22	Isoform 1 of Heat shock cognate 71 kDa protein
IPI00419373	2	Isoform 1 of Heterogeneous nuclear ribonucleoprotein A3
IPI00419373	6	Isoform 1 of Heterogeneous nuclear ribonucleoprotein A3
IPI00028888	6	Isoform 1 of Heterogeneous nuclear ribonucleoprotein D0
IPI00028888	5	Isoform 1 of Heterogeneous nuclear ribonucleoprotein D0
IPI00216049	4	Isoform 1 of Heterogeneous nuclear ribonucleoprotein K
IPI00216049	5	Isoform 1 of Heterogeneous nuclear ribonucleoprotein K
IPI00216049	7	Isoform 1 of Heterogeneous nuclear ribonucleoprotein K
IPI00018140	1	Isoform 1 of Heterogeneous nuclear ribonucleoprotein Q
IPI00006187	4	Isoform 1 of Histone deacetylase 3
IPI00096972	7	Isoform 1 of Histone-lysine N-methyltransferase, H3 lysine-9 specific 3
IPI00015526	6	Isoform 1 of Histone-lysine N-methyltransferase, H3 lysine-9 specific 5
IPI00015526	5	Isoform 1 of Histone-lysine N-methyltransferase, H3 lysine-9 specific 5
IPI00296432	6	Isoform 1 of IWS1 homolog
IPI00185919	6	Isoform 1 of La-related protein 1
IPI00185919	5	Isoform 1 of La-related protein 1
IPI00185919	19	Isoform 1 of La-related protein 1
IPI00641990	5	Isoform 1 of LAS1-like protein
IPI00641990	7	Isoform 1 of LAS1-like protein
IPI00010590	18	Isoform 1 of Lymphoid-specific helicase

IPI00165357	1	Isoform 1 of Metastasis-associated protein MTA3
IPI00165357	8	Isoform 1 of Metastasis-associated protein MTA3
IPI00438701	7	Isoform 1 of Methyl-CpG-binding domain protein 1
IPI00169400	3	Isoform 1 of Mitochondrial 28S ribosomal protein S5
IPI00169400	8	Isoform 1 of Mitochondrial 28S ribosomal protein S5
IPI00023334	5	Isoform 1 of Mitochondrial 39S ribosomal protein L4
IPI00289344	13	Isoform 1 of Nuclear receptor corepressor 1
IPI00015953	8	Isoform 1 of Nucleolar RNA helicase 2
IPI00015953	7	Isoform 1 of Nucleolar RNA helicase 2
IPI00015953	11	Isoform 1 of Nucleolar RNA helicase 2
IPI00015953	19	Isoform 1 of Nucleolar RNA helicase 2
IPI00410693	5	Isoform 1 of Plasminogen activator inhibitor 1 RNA-binding protein
IPI00008524	5	Isoform 1 of Polyadenylate-binding protein 1
IPI00008524	4	Isoform 1 of Polyadenylate-binding protein 1
IPI00008524	4	Isoform 1 of Polyadenylate-binding protein 1
IPI00012726	1	Isoform 1 of Polyadenylate-binding protein 4
IPI00216046	7	Isoform 1 of Probable global transcription activator SNF2L1
IPI00216046	6	Isoform 1 of Probable global transcription activator SNF2L1
IPI00216046	4	Isoform 1 of Probable global transcription activator SNF2L1
IPI00022613	6	Isoform 1 of Probable nucleolar complex protein 14
IPI00006702	5	Isoform 1 of Proline-, glutamic acid- and leucine-rich protein 1
IPI00006702	5	Isoform 1 of Proline-, glutamic acid- and leucine-rich protein 1
IPI00101186	5	Isoform 1 of Protein KIAA0690
IPI00101186	15	Isoform 1 of Protein KIAA0690
IPI00072377	8	Isoform 1 of Protein SET
IPI00072377	10	Isoform 1 of Protein SET
IPI00072377	8	Isoform 1 of Protein SET
IPI00411733	7	Isoform 1 of Putative ATP-dependent RNA helicase DHX30
IPI00006932	2	Isoform 1 of Putative RNA-binding protein Luc7-like 2
IPI00375358	5	Isoform 1 of Replication factor C subunit 1
IPI00375358	3	Isoform 1 of Replication factor C subunit 1
IPI00375358	4	Isoform 1 of Replication factor C subunit 1
IPI00163505	6	Isoform 1 of RNA-binding protein 39
IPI00163505	3	Isoform 1 of RNA-binding protein 39
IPI00163505	5	Isoform 1 of RNA-binding protein 39
IPI00163505	6	Isoform 1 of RNA-binding protein 39
IPI00163505	10	Isoform 1 of RNA-binding protein 39
IPI00021187	5	Isoform 1 of RuvB-like 1
IPI00021187	4	Isoform 1 of RuvB-like 1
IPI00021187	7	Isoform 1 of RuvB-like 1
IPI00021187	9	Isoform 1 of RuvB-like 1
IPI00745872	7	Isoform 1 of Serum albumin precursor
IPI00478292	38	Isoform 1 of Spectrin alpha chain, brain
IPI00478292	23	Isoform 1 of Spectrin alpha chain, brain
IPI00478292	44	Isoform 1 of Spectrin alpha chain, brain
IPI00478292	7	Isoform 1 of Spectrin alpha chain, brain
IPI00478292	10	Isoform 1 of Spectrin alpha chain, brain
IPI00149044	8	Isoform 1 of Suppressor of hairy wing homolog 4

IPI00339381	5	Isoform 1 of SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3
IPI00216047	2	Isoform 1 of SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member 2
IPI00062599	3	Isoform 1 of SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1
IPI00298058	5	Isoform 1 of Transcription elongation factor SPT5
IPI00159322	7	Isoform 1 of Transcription factor 20
IPI00159322	7	Isoform 1 of Transcription factor 20
IPI00220107	3	Isoform 1 of Transcriptional regulator ATRX
IPI00410330	2	Isoform 1 of Transcriptional repressor p66 alpha
IPI00410330	9	Isoform 1 of Transcriptional repressor p66 alpha
IPI00107113	6	Isoform 1 of U3 small nucleolar RNA-associated protein 14 homolog A
IPI00292000	10	Isoform 1 of U4/U6 small nuclear ribonucleoprotein Prp31
IPI00292000	5	Isoform 1 of U4/U6 small nuclear ribonucleoprotein Prp31
IPI00292000	11	Isoform 1 of U4/U6 small nuclear ribonucleoprotein Prp31
IPI00292000	6	Isoform 1 of U4/U6 small nuclear ribonucleoprotein Prp31
IPI00292000	6	Isoform 1 of U4/U6 small nuclear ribonucleoprotein Prp31
IPI00000728	5	Isoform 1 of Ubiquitin carboxyl-terminal hydrolase 15
IPI00410067	5	Isoform 1 of Zinc finger CCCH type antiviral protein 1
IPI00029484	6	Isoform 1 of Zinc finger MYM-type protein 3
IPI00029484	11	Isoform 1 of Zinc finger MYM-type protein 3
IPI00003886	3	Isoform 2 of Guanine nucleotide-binding protein-like 3
IPI00003886	6	Isoform 2 of Guanine nucleotide-binding protein-like 3
IPI00155577	5	Isoform 2 of HEAT repeat-containing protein 3
IPI00216746	4	Isoform 2 of Heterogeneous nuclear ribonucleoprotein K
IPI00217540	6	Isoform 2 of Lysine-specific histone demethylase 1
IPI00084571	10	Isoform 2 of Mitochondrial 39S ribosomal protein L39
IPI00026466	14	Isoform 2 of Nipped-B-like protein
IPI00220740	8	Isoform 2 of Nucleophosmin
IPI00220740	3	Isoform 2 of Nucleophosmin
IPI00220740	10	Isoform 2 of Nucleophosmin
IPI00150057	5	Isoform 2 of SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member 2
IPI00402183	3	Isoform 3 of Heterogeneous nuclear ribonucleoprotein Q
IPI00005780	6	Isoform 3 of UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit
IPI00005780	6	Isoform 3 of UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit
IPI00106509	2	Isoform 4 of Heterogeneous nuclear ribonucleoprotein A/B
IPI00106509	3	Isoform 4 of Heterogeneous nuclear ribonucleoprotein A/B
IPI00106509	3	Isoform 4 of Heterogeneous nuclear ribonucleoprotein A/B
IPI00106509	4	Isoform 4 of Heterogeneous nuclear ribonucleoprotein A/B
IPI00219330	3	Isoform 5 of Interleukin enhancer-binding factor 3
IPI00219330	4	Isoform 5 of Interleukin enhancer-binding factor 3
IPI00219330	5	Isoform 5 of Interleukin enhancer-binding factor 3
IPI00219330	5	Isoform 5 of Interleukin enhancer-binding factor 3
IPI00023097	9	Isoform 5 of Protein polybromo-1
IPI00023097	7	Isoform 5 of Protein polybromo-1
IPI00023097	8	Isoform 5 of Protein polybromo-1
IPI00396378	3	Isoform B1 of Heterogeneous nuclear ribonucleoproteins A2/B1
IPI00396378	5	Isoform B1 of Heterogeneous nuclear ribonucleoproteins A2/B1
IPI00396378	6	Isoform B1 of Heterogeneous nuclear ribonucleoproteins A2/B1
IPI00396378	11	Isoform B1 of Heterogeneous nuclear ribonucleoproteins A2/B1

IPI00396378	14	Isoform B1 of Heterogeneous nuclear ribonucleoproteins A2/B1
IPI00216592	6	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2
IPI00216592	12	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2
IPI00216592	11	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2
IPI00216592	10	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2
IPI00477313	1	Isoform C2 of Heterogeneous nuclear ribonucleoproteins C1/C2
IPI00013933	9	Isoform DPI of Desmoplakin
IPI00384456	10	Isoform GTBP-N of DNA mismatch repair protein MSH6
IPI00384456	8	Isoform GTBP-N of DNA mismatch repair protein MSH6
IPI00012773	6	Isoform Long of Metastasis-associated protein MTA1
IPI00012773	7	Isoform Long of Metastasis-associated protein MTA1
IPI00012773	18	Isoform Long of Metastasis-associated protein MTA1
IPI00005614	31	Isoform Long of Spectrin beta chain, brain 1
IPI00005614	29	Isoform Long of Spectrin beta chain, brain 1
IPI00005614	38	Isoform Long of Spectrin beta chain, brain 1
IPI00005614	11	Isoform Long of Spectrin beta chain, brain 1
IPI00005184	12	Isoform Long of Transcription intermediary factor 1-alpha
IPI00177817	2	Isoform SERCA2A of Sarcoplasmic/endoplasmic reticulum calcium ATPase 2
IPI00328230	9	Isoform Short of Spectrin beta chain, brain 1
IPI00554711	6	Junction plakoglobin
IPI00789324	8	JUP protein
IPI00292715	2	keratin 34
IPI00290078	2	keratin 4
IPI00166205	6	Keratin 5b
IPI00174775	1	Keratin 6 irs3
IPI00816709	1	Keratin 6C
IPI00376379	1	Keratin 77
IPI00182654	1	Keratin, hair, basic, 1
IPI00009865	1	Keratin, type I cytoskeletal 10
IPI00384444	17	Keratin, type I cytoskeletal 14
IPI00384444	3	Keratin, type I cytoskeletal 14
IPI00384444	14	Keratin, type I cytoskeletal 14
IPI00384444	3	Keratin, type I cytoskeletal 14
IPI00217963	7	Keratin, type I cytoskeletal 16
IPI00217963	2	Keratin, type I cytoskeletal 16
IPI00217963	3	Keratin, type I cytoskeletal 16
IPI00217963	7	Keratin, type I cytoskeletal 16
IPI00450768	2	Keratin, type I cytoskeletal 17
IPI00019359	27	Keratin, type I cytoskeletal 9
IPI00019359	19	Keratin, type I cytoskeletal 9
IPI00019359	10	Keratin, type I cytoskeletal 9
IPI00019359	16	Keratin, type I cytoskeletal 9
IPI00220327	29	Keratin, type II cytoskeletal 1
IPI00220327	23	Keratin, type II cytoskeletal 1
IPI00220327	22	Keratin, type II cytoskeletal 1
IPI00021304	17	Keratin, type II cytoskeletal 2 epidermal
IPI00021304	22	Keratin, type II cytoskeletal 2 epidermal
IPI00021304	11	Keratin, type II cytoskeletal 2 epidermal

IPI00021304	7	Keratin, type II cytoskeletal 2 epidermal
IPI00008359	3	Keratin, type II cytoskeletal 2 oral
IPI00290857	1	Keratin, type II cytoskeletal 3
IPI00009867	13	Keratin, type II cytoskeletal 5
IPI00009867	4	Keratin, type II cytoskeletal 5
IPI00009867	12	Keratin, type II cytoskeletal 5
IPI00293665	7	Keratin, type II cytoskeletal 6B
IPI00293665	5	Keratin, type II cytoskeletal 6B
IPI00299145	7	Keratin, type II cytoskeletal 6E
IPI00554648	2	Keratin, type II cytoskeletal 8
IPI00554648	3	Keratin, type II cytoskeletal 8
IPI00514908	4	Keratinocyte proline-rich protein
IPI00175151	3	KIAA1546 protein (Fragment)
IPI00305289	13	Kinesin-like protein KIF11
IPI00305289	4	Kinesin-like protein KIF11
IPI00305289	15	Kinesin-like protein KIF11
IPI00029081	10	ligase III, DNA, ATP-dependent isoform alpha precursor
IPI00031397	4	Long-chain-fatty-acid--CoA ligase 3
IPI00153032	4	LTV1 homolog
IPI00641950	5	Lung cancer oncogene 7
IPI00641950	7	Lung cancer oncogene 7
IPI00010388	4	Major centromere autoantigen B
IPI00017297	4	Matrin-3
IPI00017297	4	Matrin-3
IPI00017297	5	Matrin-3
IPI00017297	2	Matrin-3
IPI00017297	3	Matrin-3
IPI00171798	10	Metastasis-associated protein MTA2
IPI00171798	12	Metastasis-associated protein MTA2
IPI00418234	7	Methyl-CpG-binding protein 2
IPI00012202	6	Methylosome protein 50
IPI00012202	7	Methylosome protein 50
IPI00004795	2	Methylosome subunit pICln
IPI00013146	10	Mitochondrial 28S ribosomal protein S22
IPI00013146	13	Mitochondrial 28S ribosomal protein S22
IPI00022002	10	Mitochondrial 28S ribosomal protein S27
IPI00022002	13	Mitochondrial 28S ribosomal protein S27
IPI00018120	12	Mitochondrial 28S ribosomal protein S29
IPI00018120	14	Mitochondrial 28S ribosomal protein S29
IPI00010278	6	Mitochondrial 28S ribosomal protein S30
IPI00169413	5	Mitochondrial 28S ribosomal protein S34
IPI00549381	5	Mitochondrial ribosoMal protein L1
IPI00032881	4	Mitochondrial ribosomal protein S23
IPI00006440	7	mitochondrial ribosomal protein S7
IPI00014835	10	mitochondrial ribosomal protein S9
IPI00014835	11	mitochondrial ribosomal protein S9
IPI00019733	6	mRNA-associated protein mrnp 41
IPI00019502	47	Myosin-9

IPI00019502	7	Myosin-9
IPI00019380	3	Nuclear cap-binding protein subunit 1
IPI00290562	3	nuclear factor, interleukin 3 regulated
IPI00290562	3	nuclear factor, interleukin 3 regulated
IPI00031812	5	Nuclease sensitive element-binding protein 1
IPI00031812	8	Nuclease sensitive element-binding protein 1
IPI00015808	5	Nucleolar GTP-binding protein 2
IPI00022970	3	Nucleoprotein TPR
IPI00022970	8	Nucleoprotein TPR
IPI00023860	7	Nucleosome assembly protein 1-like 1
IPI00023860	8	Nucleosome assembly protein 1-like 1
IPI00023860	12	Nucleosome assembly protein 1-like 1
IPI00017763	3	Nucleosome assembly protein 1-like 4
IPI00382617	5	P37 AUF1
IPI00170596	8	Paired amphipathic helix protein Sin3a
IPI00300659	16	Parafibromin
IPI00103525	3	paraspeckle protein 1
IPI00300333	11	PD2 protein
IPI00783302	10	Pentatricopeptide repeat domain 3
IPI00454969	9	PHD finger protein 14
IPI00454969	6	PHD finger protein 14
IPI00472782	9	PHD finger protein 14 isoform 1
IPI00291916	3	pleckstrin homology domain interacting protein
IPI00000495	4	PNAS-125
IPI00410718	17	pogo transposable element with ZNF domain isoform 2
IPI00410718	8	pogo transposable element with ZNF domain isoform 2
IPI00449049	26	Poly [ADP-ribose] polymerase 1
IPI00449049	12	Poly [ADP-ribose] polymerase 1
IPI00449049	32	Poly [ADP-ribose] polymerase 1
IPI00016610	2	Poly(rC)-binding protein 1
IPI00016610	7	Poly(rC)-binding protein 1
IPI00016610	5	Poly(rC)-binding protein 1
IPI00016610	6	Poly(rC)-binding protein 1
IPI00299526	8	Polycomb protein SUZ12
IPI00299526	5	Polycomb protein SUZ12
IPI00004968	3	Pre-mRNA-processing factor 19
IPI00305068	7	Pre-mRNA-processing factor 6
IPI00305068	5	Pre-mRNA-processing factor 6
IPI00007928	7	Pre-mRNA-processing-splicing factor 8
IPI00007928	4	Pre-mRNA-processing-splicing factor 8
IPI00007928	6	Pre-mRNA-processing-splicing factor 8
IPI00007928	6	Pre-mRNA-processing-splicing factor 8
IPI00006725	1	Probable ATP-dependent RNA helicase DDX23
IPI00329791	9	Probable ATP-dependent RNA helicase DDX46
IPI00017617	1	Probable ATP-dependent RNA helicase DDX5
IPI00017617	4	Probable ATP-dependent RNA helicase DDX5
IPI00004459	4	Probable dimethyladenosine transferase
IPI00217055	3	Prolyl 3-hydroxylase 2 precursor

IPI00105598	2	Proteasome 26S non-ATPase subunit 11 variant (Fragment)
IPI00441473	15	Protein arginine N-methyltransferase 5
IPI00441473	10	Protein arginine N-methyltransferase 5
IPI00441473	11	Protein arginine N-methyltransferase 5
IPI00441473	8	Protein arginine N-methyltransferase 5
IPI00441473	11	Protein arginine N-methyltransferase 5
IPI00300376	3	Protein-glutamine gamma-glutamyltransferase E precursor
IPI00217413	6	Putative ATP-dependent RNA helicase DHX29
IPI00217686	5	Putative rRNA methyltransferase 3
IPI00479186	5	pyruvate kinase 3 isoform 1
IPI00479186	15	pyruvate kinase 3 isoform 1
IPI00418316	4	RACK7 isoform e
IPI00418316	6	RACK7 isoform e
IPI00001661	4	regulator of chromosome condensation 1 isoform a
IPI00001661	6	regulator of chromosome condensation 1 isoform a
IPI00290652	17	remodeling and spacing factor 1
IPI00290652	4	remodeling and spacing factor 1
IPI00290652	12	remodeling and spacing factor 1
IPI00290652	14	remodeling and spacing factor 1
IPI00006970	5	Ribosomal protein S2, bacterial and organelle form family protein
IPI00413108	4	Ribosomal protein SA
IPI00413108	9	Ribosomal protein SA
IPI00413108	4	Ribosomal protein SA
IPI00413108	13	Ribosomal protein SA
IPI00005036	3	RNA-binding protein 5
IPI00005036	4	RNA-binding protein 5
IPI00009104	5	RuvB-like 2
IPI00009104	4	RuvB-like 2
IPI00009104	6	RuvB-like 2
IPI00009104	7	RuvB-like 2
IPI00009104	6	RuvB-like 2
IPI00295940	12	Sad1/unc-84-like protein 2
IPI00295940	9	Sad1/unc-84-like protein 2
IPI00295940	7	Sad1/unc-84-like protein 2
IPI00295940	6	Sad1/unc-84-like protein 2
IPI00300786	2	Salivary alpha-amylase precursor
IPI00456887	5	Scaffold attachment factor A2
IPI00456887	5	Scaffold attachment factor A2
IPI00456887	6	Scaffold attachment factor A2
IPI00456887	6	Scaffold attachment factor A2
IPI00456887	5	Scaffold attachment factor A2
IPI00027251	11	Serine/threonine-protein kinase 38
IPI00027251	9	Serine/threonine-protein kinase 38
IPI00027251	17	Serine/threonine-protein kinase 38
IPI00237011	6	Serine/threonine-protein kinase 38-like
IPI00237011	9	Serine/threonine-protein kinase 38-like
IPI00180615	4	Similar to E2F transcription factor 7
IPI00737498	5	similar to MAX-interacting protein isoform 6

IPI00737498	6	similar to MAX-interacting protein isoform 6
IPI00143753	7	similar to U2-associated SR140 protein
IPI00143753	5	similar to U2-associated SR140 protein
IPI00143753	8	similar to U2-associated SR140 protein
IPI00029822	3	SMARCA4 isoform 2
IPI00029822	4	SMARCA4 isoform 2
IPI00029822	16	SMARCA4 isoform 2
IPI00029822	11	SMARCA4 isoform 2
IPI00333015	1	spectrin, beta, non-erythrocytic 1 isoform 2
IPI00333015	1	spectrin, beta, non-erythrocytic 1 isoform 2
IPI00550655	3	Spindlin-1
IPI00017451	4	Splicing factor 3 subunit 1
IPI00029764	6	Splicing factor 3A subunit 3
IPI00029764	4	Splicing factor 3A subunit 3
IPI00026089	19	Splicing factor 3B subunit 1
IPI00026089	16	Splicing factor 3B subunit 1
IPI00026089	26	Splicing factor 3B subunit 1
IPI00026089	19	Splicing factor 3B subunit 1
IPI00026089	12	Splicing factor 3B subunit 1
IPI00221106	7	splicing factor 3B subunit 2
IPI00221106	10	splicing factor 3B subunit 2
IPI00221106	12	splicing factor 3B subunit 2
IPI00300371	20	Splicing factor 3B subunit 3
IPI00300371	14	Splicing factor 3B subunit 3
IPI00300371	23	Splicing factor 3B subunit 3
IPI00300371	26	Splicing factor 3B subunit 3
IPI00300371	14	Splicing factor 3B subunit 3
IPI00176706	5	Splicing factor 45
IPI00176706	5	Splicing factor 45
IPI00031556	7	Splicing factor U2AF 65 kDa subunit
IPI00031556	4	Splicing factor U2AF 65 kDa subunit
IPI00031556	4	Splicing factor U2AF 65 kDa subunit
IPI00031556	7	Splicing factor U2AF 65 kDa subunit
IPI00010204	2	Splicing factor, arginine/serine-rich 3
IPI00007765	6	Stress-70 protein, mitochondrial precursor
IPI00007765	3	Stress-70 protein, mitochondrial precursor
IPI00007765	6	Stress-70 protein, mitochondrial precursor
IPI00183642	3	Suppressor of hairy wing homolog 3
IPI00297211	30	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5
IPI00297211	9	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5
IPI00297211	29	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5
IPI00297211	18	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5
IPI00234252	4	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member 1
IPI00234252	6	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member 1
IPI00234252	10	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member 1
IPI00234252	6	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member 1
IPI00025815	2	TAR DNA-binding protein 43
IPI00025815	2	TAR DNA-binding protein 43

IPI00549664	3	Testis-expressed sequence 10 protein
IPI00549664	5	Testis-expressed sequence 10 protein
IPI00104050	2	Thyroid hormone receptor-associated protein 3
IPI00104050	3	Thyroid hormone receptor-associated protein 3
IPI00413755	1	Transcription initiation factor TFIID subunit 4
IPI00103554	4	Transcriptional repressor p66 beta
IPI00103554	5	Transcriptional repressor p66 beta
IPI00103554	17	Transcriptional repressor p66 beta
IPI00022774	11	Transitional endoplasmic reticulum ATPase
IPI00022774	5	Transitional endoplasmic reticulum ATPase
IPI00022774	5	Transitional endoplasmic reticulum ATPase
IPI00438229	5	tripartite motif-containing 28 protein
IPI00438229	11	tripartite motif-containing 28 protein
IPI00438229	7	tripartite motif-containing 28 protein
IPI00438229	9	tripartite motif-containing 28 protein
IPI00438229	13	tripartite motif-containing 28 protein
IPI00010948	4	Tripartite motif-containing protein 26
IPI00292894	12	TSR1, 20S rRNA accumulation, homolog
IPI00166768	4	TUBA6 protein
IPI00007750	2	Tubulin alpha-1 chain
IPI00007750	1	Tubulin alpha-1 chain
IPI00180675	6	Tubulin alpha-3 chain
IPI00180675	1	Tubulin alpha-3 chain
IPI00180675	1	Tubulin alpha-3 chain
IPI00218343	12	Tubulin alpha-6 chain
IPI00387144	6	Tubulin alpha-ubiquitous chain
IPI00387144	6	Tubulin alpha-ubiquitous chain
IPI00011654	2	Tubulin beta chain
IPI00011654	7	Tubulin beta chain
IPI00011654	2	Tubulin beta chain
IPI00011654	2	Tubulin beta chain
IPI00011654	5	Tubulin beta chain
IPI00013475	1	Tubulin beta-2A chain
IPI00007752	10	Tubulin beta-2C chain
IPI00007752	1	Tubulin beta-2C chain
IPI00007752	17	Tubulin beta-2C chain
IPI00007752	9	Tubulin beta-2C chain
IPI00375911	1	Type I inner root sheath specific keratin 25 irs1
IPI00021417	3	U4/U6.U5 tri-snRNP-associated protein 1
IPI00420014	10	U5 small nuclear ribonucleoprotein 200 kDa helicase
IPI00420014	10	U5 small nuclear ribonucleoprotein 200 kDa helicase
IPI00026320	2	Ubiquitin-protein ligase EDD1
IPI00385928	2	Uncharacterized protein C4orf14
IPI00385928	6	Uncharacterized protein C4orf14
IPI00418471	8	Vimentin
IPI00418471	5	Vimentin
IPI00418471	4	Vimentin
IPI00005492	3	WD repeat protein 5

IPI00005492	3	WD repeat protein 5
IPI00005492	3	WD repeat protein 5
IPI00019269	4	WD repeat protein 61
IPI00016542	3	WD repeat protein 76
IPI00477949	3	zinc finger protein 262
IPI00022460	3	Zinc finger protein 592
IPI00064212	5	Zinc finger protein KIAA1802
IPI00064212	6	Zinc finger protein KIAA1802

References

Allan AM, Liang X, Luo Y, Pak C, Li X, Szulwach KE, Chen D, Jin P, Zhao X. (2008). The loss of methyl-CpG binding protein 1 leads to autism-like behavioral deficits. Hum Mol Genet. 17, 2047-2057

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. (1999). Rett syndrome is caused by mutations in X-linked MeCP2, encoding methyl-CpG binding protein 2. Nat Genet. 23, 185-8

Antequera F, Boyes J, Bird A. (1990). High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell. 62, 503-14

Antequera F, Macleod D, Bird AP. (1989). Specific protection of methylated CpGs in mammalian nuclei. Cell. 58, 509-17

Avery OT, Macleod CM, McCarty M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J. Exp. Med. 79, 137-158

Baba D, Maita N, Jee JG, Uchimura Y, Saitoh H, Sugasawa K, Hanaoka F, Hiroaki H, Shirakawa M. (2005). Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. Nature. 435, 979-82

Bárdos JI, Saurin AJ, Tissot C, Duprez E, Freemont PS. (2000). HPC3 is a new human polycomb orthologue that interacts and associates with RING1 and Bmi1 and has transcriptional repression properties. J Biol Chem. 275, 28785-92

Barker D, Schafer M, White R. (1984). Restriction sites containing CpG show a high frequency of polymorphism in human DNA. Cell.36, 131-138

Barr H, Hermann A, Berger J, Tsai HH, Adie K, Prokhortchouk A, Hendrich B, Bird A. (2007). Mbd2 contributes to DNA methylation-directed repression of the Xist gene. Mol Cell Biol. 27, 3750-7

Bartolomei MS, Webber AL, Brunkow ME, Tilghman SM. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. Genes Dev. 7, 1663-73

Bartolomei MS, Zemel S, Tilghman SM. (1991). Parental imprinting of the mouse H19 gene. Nature. 351, 153-5

Beard C, Li E, Jaenisch R. (1995). Loss of methylation activates Xist in somatic but not in embryonic cells. Genes Dev. 9, 2325-34

Bell AC, Felsenfeld G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. 405, 482-5

Ben-Hattar J, Jiricny J. (1988). Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo. Gene. 65, 219-27

Berger J, Sansom O, Clarke A, Bird A. (2007). MBD2 is required for correct spatial gene expression in the gut. Mol Cell Biol. 27, 4049-57

Bestor TH, Ingram VM. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity and mode of interaction with DNA. Proc Natl Acad Sci U S A. 80, 5559-63

Bestor T, Laudano A, Mattaliano R, Ingram V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol. 203, 971-83

Bird A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. 16, 6-21

Bird AP. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8, 1499-1504

Bird A. (1997). Does DNA methylation control transposition of selfish elements in the germline? Trends Genet. 13, 469-70

Bird AP. (1978). Use of Restriction Enzymes to Study Eukaryotic DNA Methylation: II. The Symmetry of Methylated Sites Supports Semi-conservative Copying of the Methylation Pattern. J. Mol. Biol. 118, 49-60.

Bird AP, Taggart MH, Smith BA. (1979). Methylated and Unmethylated DNA Compartments in the Sea Urchin Genome. Cell. 17, 889-901.

Bird AP, Taggart MH. (1980). Variable patterns of total DNA and rDNA methylation in animals. Nucleic Acids Res. 8, 1485-1497.

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. (1985). A Fraction of the Mouse Genome That is Derived from Islands of Nonmethylated, CpG-Rich DNA. Cell. 40, 91-99

Brackertz M, Gong Z, Leers J, Renkawitz R. (2006). p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. Nucleic Acids Res. 34, 397-406

Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H. (1994). Sp1 elements protect a CpG island from de novo methylation. Nature. 371, 435-438

Brès V, Yoh SM, Jones KA. (2008). The multi-tasking P-TEFb complex. Curr Opin Cell Biol. 20, 334-40

Bochar DA, Savard J, Wang W, Lafleur DW, Moore P, Côté J, Shiekhattar R. (2000). A family of chromatin remodeling factors related to Williams syndrome transcription factor. Proc Natl Acad Sci U S A. 97, 1038-43

Bostick M, Kim JK, Estève PO, Clark A, Pradhan S, Jacobsen SE. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 317, 1760-4

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science. 294, 2536-9

Bourc'his D, Bestor TH. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 431, 96-9

Boyes J, Bird A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell. 64, 1123-34

Burdon RH, Adams RL. (1969). The in vivo methylation of DNA in mouse fibroblasts. Biochim Biophys Acta. 174, 322-9

Buschhausen G, Graessmann M, Graessmann A. (1985). Inhibition of herpes simplex thymidine kinase gene expression by DNA methylation is an indirect effect. Nucleic Acids Res. 13, 5503-13

Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M, Jacobsen SE. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Curr Biol. 13, 2212-7

Carrel L, Willard HF. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature. 434, 400-4

Cedar H, Solage A, Glaser G, Razin A. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. Nucleic Acids Res. 6, 2125-2132

Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J, Zoghbi HY. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. Science. 320, 1224-9

Chan MF, van Amerongen R, Nijjar T, Cuppen E, Jones PA, Laird PW. (2001). Reduced rates of gene loss, gene silencing, and gene mutation in Dnmt1-deficient embryonic stem cells. Mol Cell Biol. 21, 7587-600

Chan SW, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE. (2004). RNA silencing genes control de novo DNA methylation. Science. 303, 1336

Chang Q, Khare G, Dani V, Nelson S, Jaenisch R. (2006). The disease progression of Mecp2 mutant mice is affected by the level of BDNF expression. Neuron. 49, 341-8

Chaudhary K, Deb S, Moniaux N, Ponnusamy MP, Batra SK. (2007) Human RNA polymerase II-associated factor complex: dysregulation in cancer. Oncogene. 26, 7499-507

Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. (1998). DNA hypomethylation leads to elevated mutation rates. Nature. 395, 89-93

Chen RZ, Akbarian S, Tudor M, Jaenisch R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat Genet. 27, 327-31

Chen T, Ueda Y, Dodge JE, Wang Z, Li E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol. 23, 5594-605

Chen WG, Chang Q, Lin Y, Meissner A, West AE, Griffith EC, Jaenisch R, Greenberg ME. (2003). Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science. 302, 885-9

Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev. 23, 105-17

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science. 277, 1996-2000

Cooper DN, Taggart MH, Bird AP. (1983). Unmethylated domains in vertebrate DNA. Nucleic Acids Res. 11, 647-658.

Cooper DN, Youssoufian H. (1988). The CpG dinucleotides and human genetic disease. Hum Genet. 78, 151-5

Cross SH, Meehan RR, Nan X, Bird A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet. 16, 256-259

Damelin M, Bestor TH. (2007). Biological Functions of DNA Methyltransferase 1 Require Its Methyltransferase Activity. Mol Cell Biol. 27, 3891-9

Daniel JM, Spring CM, Crawford HC, Reynolds AB, Baig A. (2002). The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence specific consensus and methylated CpG dinucleotides. Nucleic Acids Res. 30, 2911-9

DeChiara TM, Robertson EJ, Efstratiadis A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. Cell. 64, 849-59

Dennis K, Fan T, Geiman T, Yan Q, Muegge K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev. 15, 2940-4

Desterro JM, Rodriguez MS, Hay RT. (1998). SUMO-1 modification of IkappaBalpha inhibits NF-kappaB activation. Mol Cell. 2, 233-9

Dorsett D. (1993). Distance-independent inactivation of an enhancer by the suppressor of Hairy-wing DNA-binding protein of Drosophila. Genetics. 134, 1135-44

Fan T, Hagan JP, Kozlov SV, Stewart CL, Muegge K. (2005). Lsh controls silencing of the imprinted Cdkn1c gene. Development. 132, 635-44

Feinberg AP, Vogelstein B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 301, 89-92

Feldman N, Gerson A, Fang J, Li E, Zhang Y, Shinkai Y, Cedar H, Bergman Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol. 8, 188-94

Feng Q, Zhang Y. (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev. 15, 827-32

Ferguson-Smith AC, Sasaki H, Cattanach BM, Surani MA. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. Nature. 362, 751-5

Filion GJ, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez PA. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol. 26, 169-81

Foss HM, Roberts CJ, Claeys KM, Selker EU. (1993). Abnormal chromosome behavior in Neurospora mutants defective in DNA methylation. Science. 262, 1737-41

Freitag M, Hickey PC, Khlafallah TK, Read ND, Selker EU. (2004). HP1 is essential for DNA methylation in neurospora. Mol Cell. 13, 427-34

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A. 89, 1827-31

Fujita N, Takebayashi SI, Okumura K, Kudo S, Chiba T, Saya H, Nakao M. (1999). Methylation-Mediated Transcriptional Silencing in Euchromatin by Methyl-CpG Binding Protein MBD1 Isoforms. Mol Cell Biol. 19, 6415-6426.

Fujita N, Shimotake N, Ohki I, Chiba T, Saya H, Shirakawa M, Nakao M. (2000). Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. Mol Cell Biol. 20, 5107-18.

Fujita N, Watanabe S, Ichimura T, Ohkuma Y, Chiba T, Saya H, Nakao M. (2003a). MCAF Mediates MBD1-Dependent Transcriptional Repression. Mol Cell Biol. 23, 2834-43

Fujita N, Watanabe S, Ichimura T, Tsuruzoe S, Shinkai Y, Tachibana M, Chiba T, and Nakao M. (2003b). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. J. Biol. Chem. 278, 24132-24138

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet. 24, 88-91

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. (2003). The methyl-CpG binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem. 278, 4035-40

Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW, Ehrlich M. (1983). The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. 11, 6883-94

Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R. (2003). Induction of Tumors in Mice by Genomic Hypomethylation. Science. 300, 489-92

Gazin C, Wajapeyee N, Gobeil S, Virbasius CM, Green MR. (2007). An elaborate pathway required for Ras-mediated epigenetic silencing. Nature. 449, 1073-7

Gibbons RJ, McDowell TL, Raman S, O'Rourke DM, Garrick D, Ayyub H, Higgs DR. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet. 24, 368-71

Gill G. (2005). Something about SUMO inhibits transcription. Curr Opin Genet Dev. 15, 536-41

Goll MG, Bestor TH. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem. 74, 481-514

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science. 311, 395-8

Gómez-del Arco P, Koipally J, Georgopoulos K. (2005). Ikaros SUMOylation: switching out of repression. Mol Cell Biol. 25, 2688-97

Goodson ML, Hong Y, Rogers R, Matunis MJ, Park-Sarge OK, Sarge KD. (2001). Sumo-1 modification regulates DNA binding activity of heat shock transcription factor 2, a promyelocytic leukemia nuclear body associated factor. J Biol Chem. 276, 18513-8

Goyon C, Rossignol JL, Faugeron G. (1996). Native DNA repeats and methylation in Ascobolus. Nucleic Acids Res. 24, 3348-56

Gregory RI, Randall TE, Johnson CA, Khosla S, Hatada I, O'Neill LP, Turner BM, Feil R. (2001). DNA methylation is linked to deacetylation of histone H3, but not H4, on the imprinted genes Snrpn and U2af1rs1. Mol Cell Biol. 21, 5426-36

Grippo P, Laccarino M, Parisi E, Scarano E. (1968). Methylation of DNA in developing sea urchin embryos. J Mol Biol. 36, 195-208

Gruenbaum Y, Cedar H, Razin A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. Nature. 295, 620-2

Guy J, Hendrich B, Holmes M, Martin JE, Bird A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet. 27, 322-6

Hansen RS, Gartler SM. (1990). 5-Azacytidine-induced reactivation of the human X chromosome-linked PGK1 gene is associated with a large region of cytosine demethylation in the 5' CpG island. Proc Natl Acad Sci U S A. 87, 4174-8

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A. 96, 14412-7

Harikrishnan KN, Chow MZ, Baker EK, Pal S, Bassal S, Brasacchio D, Wang L Craig JM, Jones PL, Sif S, El-Osta A. (2005). Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat Genet. 37, 254-64

Harikrishnan KN, Pal S, Yarski M, Baker EK, Chow MZ, de Silva MG, Okabe J, Wang L, Jones PL, Sif S, El-Osta A. (2006). Reply to "Testing for association between MeCP2 and the brahma-associated SWI/SNF chromatin-remodeling complex". Nat Genet. 38, 964-7

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature. 405, 486-9

Hay RT. (2005). SUMO: a history of modification. Mol Cell. 18, 1-12

Hellman A, Chess A. (2007). Gene body-specific methylation on the active X chromosome. Science. 315, 1141-3

Hendrich B, Bird A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. Mol Cell Biol. 18, 6538-6547

Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. 15, 710-23

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature. 401, 301-4

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science. 308, 118-20

Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP, Walkinshaw MD. (2008). MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. Mol Cell. 29, 525-531

Hoege C, Pfander B, Moldovan GL, Pyrowolakis G, Jentsch S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature. 419, 135-41

Holliday R, Pugh JE. (1975). DNA modification mechanisms and gene activity during development. Science. 187, 226-232

Holmstrom S, Van Antwerp ME, Iñiguez-Lluhi JA. (2003). Direct and distinguishable inhibitory roles for SUMO isoforms in the control of transcriptional synergy. Proc Natl Acad Sci U S A. 100, 15758-63

Honda S, Selker EU. (2008). Direct interaction between DNA methyltransferase DIM-2 and HP1 is required for DNA methylation in Neurospora crassa. Mol Cell Biol. 28, 6044-55

Hong Y, Rogers R, Matunis MJ, Mayhew CN, Goodson ML, Park-Sarge OK, Sarge KD. (2001). Regulation of heat shock transcription factor 1 by stress-induced SUMO-1 modification. J Biol Chem. 276, 40263-7

Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Genet. 37, 31-40

Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene. 27, 404-8

Hu K, Nan X, Bird A, Wang W. (2006). Testing for association between MeCP2 and the brahma-associated SWI/SNF chromatin-remodeling complex. Nat Genet. 38, 962-4

Hutchins AS, Mullen AC, Lee HW, Sykes KJ, High FA, Hendrich BD, Bird AP, Reiner SL. (2002). Gene silencing quantitatively controls the function of a developmental trans-activator. Mol Cell. 10, 81-91

Ichimura T, Watanabe S, Sakamoto Y, Aoto T, Fujita N, Nakao M. (2005). Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins. J Biol Chem. 280, 13928-35

Iguchi-Ariga SM, Schaffner W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes Dev. 3, 612-9

Illingworth R, Kerr A, Desousa D, Jorgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J, Humphray S, Cox T, Langford C, Bird A. (2008). PLoS Biol. 6, e22

Jackson JP, Lindroth AM, Cao X, Jacobsen SE. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. 416, 556-60

Jähner D, Stuhlmann H, Stewart CL, Harbers K, Löhler J, Simon I, Jaenisch R. (1982). De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature. 298, 623-8

Jeanpierre M, Turleau C Aurias A, Prieur M, Ledeist F, Fischer A, Viegas-Péquignot E. (1993). An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. Hum Mol Genet. 2, 731-5

Jeddeloh JA, Stokes TL, Richards EJ. (1999). Maintenance of genomic methylation requires a SWI/SNF2-like protein. Nat Genet. 22, 94-7

Jeffery L, Nakielny S. (2004). Components of the DNA methylation system of chromatin control are RNA binding proteins. 279, 49479-87

Jeltsch A. (2006). Molecular enzymology of mammalian DNA methyltransferases. Curr Top Microbiol Immunol. 301, 203-25

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo methylation. Nature. 449, 248-51

Johnson ES, Blobel G. (1997). Ubc9 is the conjugating enzyme for the ubiquitin-like protein Smt3p. J Biol Chem. 272, 26799-802

Jones BK, Levorse JM, Tilghman SM. (1998). Igf2 imprinting does not require its own DNA methylation or H19 RNA. Genes Dev. 12, 2200-7

Jones MC, Fusi L, Higham JH, Abdel-Hafiz H, Horwitz KB, Lam EW, Brosens JJ. (2006). Regulation of the SUMO pathway sensitizes differentiating human endometrial stromal cells to progesterone. Proc Natl Acad Sci U S A. 103, 16272-7

Jones PA. (1985). Altering Gene Expression with 5-Azacytidine. Cell. 40, 485-6

Jones PA, Baylin SB. (2002). The fundamental role of epigenetic events in cancer. Nat Rev Genet. 3, 415-28

Jones L, Ratcliff F, Baulcombe DC. (2001). RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. Curr Biol. 11, 747-57

Jørgensen HF, Ben-Porath I, and Bird AP. (2004). Mbd1 is Recruited to both Methylated and Nonmethylated CpGs via Distinct DNA Binding Domains. Mol Cell Biol. 24, 3387-3395

Josse J, Kaiser AD, Kornberg A. (1961). Enzymatic synthesis of deoxyribonucleic acid. VII. Frequencies of nearest neighbor base sequences in deoxyribonucleic acid. J. Biol. Chem. 236, 864-75

Jüttermann R, Li E, Jaenisch R. (1994). Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. Proc Natl Acad Sci U S A. 91, 11797-801

Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429, 900-3

Kangaspeska S, Stride B, Métivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F, Reid G. (2008). Transient cyclical methylation of promoter DNA. Nature. 452, 112-5

Kappler JW. (1970). The kinetics of DNA methylation in cultures of a mouse adrenal cell line. J Cell Physiol. 75, 21-31

Kawahara M, Wu Q, Takahashi N, Morita S, Yamada K, Ito M, Ferguson-Smith AC, Kono T. (2007). High-frequency generation of viable mice from engineered bi-maternal embryos. Nat Biotechnol. 25, 1045-50

Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, Pikarski E, Young RA, Cedar H, Simon I. (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat Genet. 38, 149-53

Klimasauskas S, Kumar S, Roberts RJ, Cheng X. (1994). HhaI Methyltransferase Flips Its Target Base Out of the DNA Helix. Cell. 76, 357-69

Klose RJ, Bird AP. (2004). MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. J. Biol. Chem. 279, 46490-46496.

Kokura K, Kaul SC, Wadhwa R, Nomura T, Khan MM, Shinagawa T, Yasukawa T, Colmenares C, Ishii S. (2001). The Ski protein family is required for MeCP2-mediated transcriptional repression. J. Biol. Chem. 276, 34115-21

Kondo E, Gu Z, Horii A, Fukushige S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. Mol Cell Biol. 25, 4388-96

Kono T, Obata Y, Wu Q, Niwa K, Ono Y, Yamamoto Y, Park ES, Seo JS, Ogawa H. (2004). Birth of parthenogenetic mice that can develop to adulthood. Nature. 428, 860-4

Kriaucionis S, Paterson A, Curtis J, Guy J, Macleod D, Bird A. (2006). Gene expression analysis exposes mitochondrial abnormalities in a mouse model of Rett syndrome. Mol Cell Biol. 26, 5033-42

Kronforst MR, Gilley DC, Strassmann JE, Queller DC. (2008). DNA methylation is widespread across social Hymenoptera. Curr Biol. 18, 287-8

Laird PW, Jackson-Grusby L, Fazeli A, Dickinson SL, Jung WE, Li E, Weinberg RA, Jaenisch R. (1995). Suppression of intestinal neoplasia by DNA hypomethylation. Cell. 81, 197-205

Lallemand-Breitenbach V, Jeanne M, Benhenda S, Nasr R, Peres L, Zhou J, Zhu J, Raught B, de Thé. (2008). Arsenic degrades PML or PML-RARalpha through a SUMO-triggered RNF4/ubiquitin-mediated pathway. Nat Cell Biol. 10, 547-55

Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WA, Cohen A, Lasonder E, Stunnenberg HG. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol. 26, 843-51

Lee DH, Goldberg AL. (1998). Proteasome inhibitors: valuable new tools for cell biologists. Trends Cell Biol. 8, 397-403

Lee MG, Norman J, Shilatifard A, Shiekhattar R. (2007). Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. Cell. 128, 877-87

Lee MG, Wynder C, Cooch N, Shiekhattar R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature. 437, 432-5

Lei H, Oh SP, Okano M, Jüttermann R, Goss KA, Jaenisch R, Li E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development. 122, 3195-205

Leighton PA, Ingram RS, Eggenschwiler J, Efstratiadis A, Tilghman SM. (1995). Disruption of imprinting caused by deletion of the H19 gene region in mice. Nature. 375, 34-9

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T , Peters AH. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol. 13, 1192-200

Leonhardt H, Page AW, Weier HU, Bestor TH. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell. 71, 865-73

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. (1992). Purification, sequence and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell. 69, 905-14

Li CF, Pontes O, El-Shami M, Henderson IR, Bernatavichute YV, Chan SW, Lagrange T, Pikaard CS, Jacobsen SE. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. Cell. 126, 93-106

Li E, Beard C, Jaenisch R. (1993). Role for DNA methylation in genomic imprinting. Nature. 366, 362-365

Li E, Bestor TH, Jaenisch R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 69, 915-26

Li H, Yamagata T, Mori M, Yasuhara A, Momoi MY. (2005). Mutation analysis of methyl-CpG binding protein family genes in autistic patients. Brain Dev. 27, 321-5

Li SJ, Hochstrasser M. (1999). A new protease required for cell-cycle progression in yeast. Nature. 398, 246-51

Li SJ, Hochstrasser M. (2003). The Ulp1 SUMO isopeptidases: distinct domains required for viability, nuclear envelope localization and substrate specificity. J Cell Biol. 160, 1069-81

Li T, Evdokimov E, Shen RF, Chao CC, Tekle E, Wang T, Stadtman ER, Yang DC, Chock PB. (2004). Sumoylation of heterogeneous nuclear ribonucleoproteins, zinc finger proteins, and nuclear pore complex proteins: a proteomic analysis. Proc Natl Acad Sci U S A. 101, 8551-6

Lima CD, Reverter D. (2008). Structure of the human SENP7 catalytic domain and poly-SUMO deconjugation activities for SENP6 and SENP7. J Biol Chem. 283, 32045-55

Lin X, Sun B, Liang M, Liang YY, Gast A, Hildebrand J, Brunicardi FC, Melchior F, Feng XH. (2003). Opposed regulation of corepressor CtBP by SUMOylation and PDZ binding. Mol Cell. 11, 1389-96

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 133, 523-36

Lock LF, Melton DW, Caskey CT, Martin GR. (1986). Methylation of the mouse hprt gene differs on the active and inactive X chromosomes. Mol Cell Biol. 6, 914-24

Lock LF, Takagi N, Martin GR. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell. 48, 39-46

Loyola A, Bonaldi T, Roche D, Imhof A, Almouzni G. (2006). PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol Cell. 24, 309-16

Lyko F, Ramsahoye BH, Jaenisch R. (2000). DNA methylation in Drosophila melanogaster. Nature. 408, 538-40

Lyon MF. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature. 190, 372-3

Lyst MJ, Nan X, Stancheva I. (2006). Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. EMBO J. 25, 5317-28

Mackay JP, Sunde M, Lowry JA, Crossley M, Matthews JM. (2007). Protein interactions: is seeing believing? Trends Biochem Sci. 32, 530-1

Macleod D, Charlton J, Mullins J, Bird AP. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. 8, 2282-92

Mahajan R, Delphin C, Guan T, Gerace L, Melchior F. (1997). A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. Cell. 88, 97-107

Mak W, Nesterova TB, de Napoles M, Appanah R, Yamanaka S, Otte AP, Brockdorff N. (2004) Reactivation of the paternal X chromosome in early mouse embryos. Science. 303, 666-9

Martinowich K, Hattori D, Wu H, Fouse S, He F, Hu Y, Fan G, Sun YE. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science. 302, 890-3

Matunis MJ, Coutavas E, Blobel G. (1996). A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. J Cell Biol. 135, 1457-70

Mayer C, Schmitz KM, Li J, Grummt I, Santoro R. (2006). Intergenic transcripts regulate the epigenetic state of rRNA genes. Mol Cell. 22, 351-61

Mayer-Jung C, Moras D, Timsit Y. (1998). Hydration and recognition of methylated CpG steps in DNA. EMBO J. 17, 2709-18

McCarty AS, Kleiger G, Eisenberg D, Smale ST. (2003). Selective dimerization of a C2H2 zinc finger subfamily. Mol Cell. 11, 459-70

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell. 58, 499-507

Meehan RR, Lewis JD, Bird AP. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res. 20, 5085-92

Melchior F, Schergaut M, Pichler A. (2003). SUMO: ligases, isopeptidases and nuclear pores. Trends Biochem Sci. 28, 612-8

Mencía MM, de Lorenzo V. (2004). Functional transplantation of the sumoylation machinery into Escherichia coli. Protein Expr Purif. 37, 409-18

Métivier R, Gallais R, Tiffoche C, Le Péron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G. (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature. 452, 45-50

Mette MF, van der Winden J, Matzke MA, Matzke AJ. (1999). Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in trans. EMBO J. 18, 241-8

Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, Keightley PD, Bishop SM, Clarke AR, Bird A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Nat Genet. 297, 403-5

Milligan L, Decourty L, Saveanu C, Rappsilber J, Ceulemans H, Jacquier A, Tollervey D. (2008). A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. Mol Cell Biol. 28, 5446-57

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature. 411, 212-4

Mohandas T, Sparkes RS, Shapiro LJ. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science. 211, 393-6

Montero LM, Filipski J, Gil P, Capel J, Martínez-Zapater JM, Salinas J. (1992). The distribution of 5-methylcytosine in the nuclear genome of plants. Nucleic Acids Res. 30, 3207-10

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr. (1996). High frequency retrotransposition in cultured mammalian cells. Cell. 87, 917-27

Morris KV, Chan SW, Jacobsen SE, Looney DJ. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science. 305, 1289-92

Mukhopadhyay D, Dasso M. (2007). Modification in reverse: the SUMO proteases. Trends Biochem Sci. 32, 286-95

Myant K, Stancheva I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. Mol Cell Biol. 28, 215-26

Nan X, Meehan RR, Bird A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res. 21, 4886-4892

Nan X, Campoy J, Bird A. (1997). MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. Cell. 88, 471-481

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature. 393, 386-9

Nan X, Hou J, Maclean A, Nasir J, Lafuente MJ, Shu X, Kriaucionis S, Bird A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proc Natl Acad Sci U S A. 104, 2709-14

Nan X, Tate P, Li E, Bird A. (1996). DNA methylation specifies chromosomal localization of MeCP2. Mol Cell Biol. 16, 414-21

Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet. 23, 58-61

Ng HH, Jeppesen P, Bird A. (2000). Active Repression of Methylated Genes by the Chromosomal Protein MBD1. Mol Cell Biol. 20, 1394-1406

Ng RK, Dean W, Dawson C, Lucifero D, Madeja Z, Reik W, Hemberger M. (2008). Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. Nat Cell Biol. 10, 1280-90

Nikitina T, Ghosh RP, Horowitz-Scherer RA, Hansen JC, Grigoryev SA, Woodcock CL. (2007). MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. J Biol Chem. 282, 28237-45

Nuber UA, Kriaucionis S, Roloff TC, Guy J, Selfridge J, Steinhoff C, Schulz R, Lipkowitz B, Ropers HH, Holmes MC, Bird A. (2005). Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. Hum Mol Genet. 14, 2247-56

Ohki I, Shimotake N, Fujita N, Nakao M, Shirakawa M. (1999). Solution structure of the methyl-CpG binding domain of the methylation-dependent transcriptional repressor MBD1. EMBO J. 18, 6653-61

Ohki I, Shimotake N, Fujita N, Jee J, Ikegami T, Nakao M, Shirakawa M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell. 105, 187-97

Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science. 303, 644-9

Okano M, Xie S, Li E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 19, 219-20

Okano M, Bell DW, Haber DA, Li E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 99, 247-57

Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell. 120, 613-22

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 448, 714-7

Palmer BR, Marinus MG. (1994). The dam and dcm strains of Escherichia coli – a review. Gene. 143, 1-12

Palmiter RD, Chen HY, Brinster RL. (1982). Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. Cell. 29, 701-10

Panning B, Jaenisch R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev. 10, 1991-2002

Panse VG, Küster B, Gerstberger T, Hurt E. (2003). Unconventional tethering of Ulp1 to the transport channel of the nuclear pore complex by karyopherins. Nat Cell Biol. 5, 21-7

Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. (1996). Requirement for Xist in X chromosome inactivation. Nature. 379, 131-7

Peters AH, Mermoud JE, O'Carroll D, Pagani M, Schweizer D, Brockdorff N, Jenuwein T. (2002). Histone H3 lysine K9 methylation is an epigenetic imprint of facultative heterochromatin. Nat Genet. 30, 77-80

Pradhan S, Bacolla A, Wells RD, Roberts RJ. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem. 274, 33002-10

Prendergast GC, Ziff EB. (1991). Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science. 251, 186-9

Proffitt JH, Davie JR, Swinton D, Hattman S. (1984). 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol. 4, 985-8

Prokhortchouk A, Hendrich B, Jørgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E. (2001). The p120 catenin Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev. 15, 1613-8

Prokhortchouk A, Sansom O, Selfridge J, Cabellero IM, Salozhin S , Aithozhina D, Cerchietti L, Meng FG, Augenlicht LH, Mariadason JM, Hendrich B, Melnick A, Prokhortchouk E, Clarke A, Bird A. (2006). Kaiso-deficient mice show resistance to intestinal cancer. Mol Cell Biol. 26, 199-208

Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ. (2006). Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature. 443, 1008-12

Rai K, Chidester S, Zavala CV, Manos EJ, James SR, Karpf AR, Jones DA, Cairns BR. (2007). Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. Genes Dev. 21, 261-6

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A. 97, 5237-42

Reese BE, Bachman KE, Baylin SB, Rountree MR. (2003). The methyl-CpG binding protein MBD1 interacts with the p150 subunit of chromatin assembly factor 1. Mol Cell Biol. 23, 3226-36

Reifsnyder C, Lowell J, Clarke A, Pillus L. (1996). Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases. Nat Genet. 14, 42-9

Reik W, Collick A, Norris ML, Barton SC, Surani MA. (1987). Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature. 328, 248-51

Reik W, Dean W, Walter J. (2001). Epigenetic reprogramming in mammalian development. Science. 293, 1089-93

Reik W, Walter J. (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet. 2, 21-32

Rideout WM 3rd, Coetzee GA, Olumi AF, Jones PA. (1990). 5-methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. Science. 249, 1288-90

Riggs AD. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet. 14, 9-25

Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. (2000). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet. 25, 338-42

Rodriguez MS, Desterro JM, Lain S, Midgley CA, Lane DP, Hay RT. (1999). SUMO-1 modification activates the transcriptional response of p53. EMBO J. 18, 6455-61

Ruzov A, Dunican DS, Prokhortchouk A, Pennings S, Stancheva I, Prokhortchouk E, Meehan RR. (2004). Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. Development. 131, 6185-94

Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E. (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev Biol. 225, 294-303

Sakamoto Y, Watanabe S, Ichimura T, Kawasuji M, Koseki H, Baba H, Nakao M. (2007). Overlapping roles of the methylated DNA-binding protein MBD1 and polycomb group proteins in transcriptional repression of HOXA genes and heterochromatin formation. J Biol Chem. 282, 16391-400

Sanders SL, Portoso M, Mata J, Bähler J, Allshire RC, Kouzarides T. (2004). Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. Cell. 119, 603-14

Santi DV, Garrett CE, Barr PJ. (1983). On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. Cell. 33, 9-10

Sapienza C, Peterson AC, Rossant J, Balling R. (1987). Degree of methylation of transgenes is dependent on gamete of origin. Nature. 328, 251-4

Sarraf SA, Stancheva I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol Cell. 15, 595-605

Scarano E, Iaccarino M, Grippo P, Winckelmans D. (1965). On methylation of DNA during development of the sea urchin embryo. J Mol Biol. 14, 603-7

Selker EU, Tountas NA, Cross SH, Margolin BS, Murphy JG, Bird AP, Freitag M. (2003). The methylated component of the Neurospora crassa genome. Nature. 422, 893-7

Seufert W, Futcher B, Jentsch S. (1995). Role of a ubiquitin-conjugating enzyme in degradation of S- and M-phase cyclins. Nature. 373, 78-81

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, Tajima S, Mitsuya K, Okano M, Koseki H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 450, 908-12

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y. (2004). Histone demethylation mediated by the nuclear oxidase homolog LSD1. Cell. 119, 941-53

Siegel LM, Monty KJ. (1966). Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. Biochem Biophys Acta. 112, 346-62

Silva AJ, Ward K, White R. (1993). Mosaic methylation in clonal tissue. Dev Biol. 156, 191-8

Simmen MW, Leitgeb S, Charlton J, Jones SJ, Harris BR, Clark VH, Bird A. (1999). Nonmethylated transposable elements and methylated genes in a chordate genome. Science. 283, 1164-7

Simpson VJ, Johnson TE, Hammen RF. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res. 14, 6711-9

Sinsheimer RL. (1955). The action of pancreatic deoxyribonuclease. II. Isomeric dinucleotides. J Biol Chem. 215, 579-83

Sneider TW, Potter VR. (1969). Methylation of mammalian DNA: studies on Novikoff hepatoma cells in tissue culture. J Mol Biol. 42, 271-84

Spada F, Haemmer A, Kuch D, Rothbauer U, Schermelleh L, Kremmer E, Carell T, Längst G, Leonhardt. (2007). DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. J Cell Biol. 176, 565-71

Stein R, Gruenbaum Y, Pollack Y, Razin A, Cedar H. (1982a). Clonal inheritance of the pattern of DNA methylation in mouse cells. Proc Natl Acad Sci U S A. 79, 61-5

Stein R, Razin A, Cedar H. (1982b). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proc Natl Acad Sci U S A. 79, 3418-22

Stewart CL, Stuhlmann H, Jähner D, Jaenisch R. (1982). De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc Natl Acad Sci U S A. 79, 4098-102

Stöger R, Kajimura TM, Brown WT, Laird CD. (1997). Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. Hum Mol Genet. 6, 1791-801

Sutter D, Doerfler W. (1980). Methylation of integrated adenovirus type 12 DNA sequences in transformed cells is inversely correlated with viral gene expression. Proc Natl Acad Sci U S A. 77, 253-6

Suzuki MM, Kerr AR, De Sousa D, Bird A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res. 17, 625-31

Swartz MN, Trautner TA, Kornberg A. (1962). Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. J. Biol. Chem. 237, 1961-7

Tachibana M, Ueda J, Fukuda M, Takeda N, Ohta T, Iwanari H, Sakihama T, Kodama T, Hamakubo T, Shinkai Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev. 19, 815-26

Takai D, Jones PA. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A. 99, 3740-5

Tamaru H, Selker EU. (2001). A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature. 414, 277-83

Tamaru H, Zhang X, McMillen D, Singh PB, Nakayama J, Grewal SI, Allis CD, Cheng X, Selker EU. (2003). Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa. Nat Genet. 34, 75-9

Tan CP, Nakielny S. (2006). Control of the DNA methylation system component MBD2 by protein arginine methylation. Mol Cell Biol. 26, 7224-35

Tatham MH, Geoffroy MC, Shen L, Plechanovova A, Hattersley N, Jaffray EG, Palvimo JJ, Hay RT. (2008). RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. Nat Cell Biol. 10, 538-46

Terpe K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol. 60, 523-33

Terui Y, Saad N, Jia S, McKeon F, Yuan J. (2004). Dual role of sumoylation in the nuclear localization and transcriptional activation of NFAT1. J Biol Chem. 279, 28257-65

Ting AH, Schuebel KE, Herman JG, Baylin SB. (2005). Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. Nat Genet. 37, 906-10

Tucker KL, Beard C, Dausmann J, Jackson-Grusby L, Laird PW, Lei H, Li E, Jaenisch R. (1996). Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. Genes Dev. 10, 1008-20

Tudor M, Akbarian S, Chen RZ, Jaenisch R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proc Natl Acad Sci U S A. 99, 15536-41

Uchimura Y, Nakamura M, Sugasawa K, Nakao M, Saitoh H. (2004). Overproduction of eukaryotic SUMO-1- and SUMO-2-conjugated proteins in Escherichia coli. Anal Biochem. 331, 204-6

Uchimura Y, Ichimura T, Uwada J, Tachibana T, Sugahara S, Nakao M, Saitoh H. (2006). Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation. J Biol Chem. 281, 23180-90

Ueba T, Kaspar B, Zhao X, Gage FH. (1999). Repression of human fibroblast growth factor 2 by a novel transcription factor. J Biol Chem. 274, 10382-7

Vanyushin BF, Tkacheva SG, Belozersky AN. (1970). Rare bases in animal DNA. Nature. 225, 948-9

Vardimon L, Kressmann A, Cedar H, Maechler M, Doerfler W. (1982). Expression of a cloned adenovirus gene is inhibited by in vitro methylation. Proc Natl Acad Sci U S A. 79, 1073-7

Vardimon L, Neumann R, Kuhlmann I, Sutter D, Doerfler W. (1980). DNA methylation and viral gene expression in adenovirus-transformed and -infected cells. Nucleic Acids Res. 8, 2461-73

Vassileva MT, Matunis MJ. (2004). SUMO modification of heterogeneous ribonucleoproteins. Mol Cell Biol. 24, 3623-32

Vertegaal AC, Ogg SC, Jaffray E, Rodriguez MS, Hay RT, Andersen JS, Mann M, Lamond AI. (2004). A proteomic study of SUMO-2 target proteins. J Biol Chem. 279, 33791-8

Viegas-Péquignot E, Dutrillaux B. (1976). Segmentation of human chromosomes induced by 5-ACR (5-azacytidine). Hum Genet. 34, 247-54

Viegas-Péquignot E, Dutrillaux B, Thomas G. (1988). Inactive X chromosome has the highest concentration of unmethylated Hha I sites. Proc Natl Acad Sci U S A. 85, 7657-60

Villa R, Morey L, Raker VA, Buschbeck M, Gutierrez A, Corsaro M, Varas F, Bossi D, Minucci S, Pelicci PG, Di Croce L. (2006). The methyl-CpG binding protein MBD1 is required for PML-RARalpha function. Proc Natl Acad Sci U S A. 103, 1400-5

Vongs A, Kakutani T, Martienssen RA, Richards EJ. (1993). Arabidopsis thaliana DNA methylation mutants. Science. 260, 1926-8

Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol. 20, 2108-21

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet. 23, 62-6

Wakefield RI, Smith BO, Nan X, Free A, Soteriou A, Uhrin D, Bird AP, Barlow PN. (1999). The solution structure of the domain from MeCP2 that binds to methylated DNA. J Mol Biol. 291, 1055-65

Walsh CP, Chaillet JR, Bestor TH. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet. 20, 116-7

Wang H, An W, Cao R, Xia L, Erdjument-Bromage H, Chatton B, Tempst P, Roeder RG, and Zhang Y. (2003). mAM Facilitates Conversion by ESET of Dimethyl to Trimethyl Lysine 9 of Histone H3 to Cause Transcriptional Repression. Mol Cell. 12, 475-487

Wang J, Hevi S, Kurash JK, Lei H, Gay F, Bajko J, Sun H, Sun W, Chang H, Xu G, Gaudet F, Li E, Chen T. (2009). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat Genet. 41, 125-9

Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA, Robinson GE. (2006). Functional CpG methylation system in a social insect. Science. 314, 645-7

Wassenegger M, Heimes S, Riedel L, Sänger HL. (1994). RNA-directed de novo methylation of genomic sequences in plants. Cell. 76, 567-76

Watanabe S, Ichimura T, Fujita N, Tsuruzoe S, Ohki I, Shirakawa M, Kawasuji M, Nakao M. (2003). Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin. Proc Natl Acad Sci U S A. 100, 12859-64.

Watson JD, Crick FHC. (1953). Molecular structure of nucleic acids. Nature. 171, 737-738

Watt F, Molloy PL. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. 2, 1136-43

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schübeler D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential methylation in normal and transformed human cells. Nat Genet. 37, 853-62

Wigler M, Levy D, Perucho M. (1981). The somatic replication of DNA methylation. Cell. 24, 33-40

Wilson DS, Keefe AD, Szostak JW. (2001). The use of mRNA display to select high-affinity protein-binding peptides. Proc Natl Acad Sci U S A. 98, 3750-55

Wilson GG, Murray NE. (1991). Restriction and modification systems. Annu Rev Genet. 25, 585-627

Wolf SF, Dintzis S, Toniolo D, Persico G, Lunnen KD, Axelman J, Migeon BR. (1984). Complete concordance between glucose-6-phosphate dehydrogenase activity and hypomethylation of 3' CpG clusters: implications for X chromosome dosage compensation. Nucleic Acids Res. 12, 9333-48

Xi S, Zhu H, Xu H, Schmidtmann A, Geiman TM, Muegge K. (2007). Lsh controls Hox gene silencing during development. Proc Natl Acad Sci U S A. 104, 14366-71

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Péquignot E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature. 402, 187-91

Yamaguchi T, Sharma P, Athanasiou M, Kumar A, Yamada S, Kuehn MR. (2005). Mutation of SENP1/SuPr-2 reveals an essential role for desumoylation in mouse development. Mol Cell Biol. 25, 5171-82

Yang SH, Jaffray E, Hay RT, Sharrocks AD. (2003). Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity. Mol Cell. 12, 63-74

Yang SH, Sharrocks AD. (2004). SUMO promotes HDAC-mediated transcriptional repression. Mol Cell. 13, 611-7

Yoder JA, Bestor TH. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet. 7, 279-84

Yoon HG, Chan DW, Reynolds AB, Qin J, Wong J. (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Mol Cell. 12, 723-34

Young JI, Hong EP, Castle JC, Crespo-Barreto J, Bowman AB, Rose MF, Kang D, Richman R, Johnson JM, Berget S, Zoghbi HY. (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. Proc Natl Acad Sci USA. 102, 17551-8

Zemel S, Bartolomei MS, Tilghman SM. (1992). Physical linkage of two mammalian imprinted genes, H19 and insulin-like growth factor 2. Nat Genet. 2, 61-5

Zhang Y, Iratni R, Erdjument-Bromage H, Tempst P, Reinberg D. (1997). Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. Cell. 89, 357-64

Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 13, 1924-35

Zhao X, Ueba T, Christie BR, Barkho B, McConnell MJ, Nakashima K, Lein ES, Eadie BD, Willhoite AR, Muotri AR, Summers RG, Chun J, Lee KF, Gage FH. (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. Proc Natl Acad Sci USA. 100, 6777-82

Zhao Y, Kwon SW, Anselmo A, Kaur K, White MA. (2004). Broad spectrum identification of cellular small ubiquitin-related modified (SUMO) substrate proteins. J Biol Chem. 279, 20999-1002

Zhu H, Geiman TM, Xi S, Jiang Q, Schmidtmann A, Chen T, Li E, Muegge K. (2006). Lsh is involved in de novo methylation of DNA. EMBO J. 25, 335-45

Zilberman D, Cao X, Johansen LK, Xie Z, Carrington JC, Jacobsen SE. (2004). Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol. 14, 1214-20

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet. 39, 61-9